

UNIVERSITÉ DE MONTRÉAL

CONTRIBUTIONS À LA MODÉLISATION DES PROCESSUS HYDROLOGIQUES

FATIMA BENSALMA
DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION
DU DIPLÔME DE PHILOSOPHIÆ DOCTOR
(MATHÉMATIQUES DE L'INGÉNIEUR)
MARS 2015

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée :

CONTRIBUTIONS À LA MODÉLISATION DES PROCESSUS HYDROLOGIQUES

présentée par : BENSALMA Fatima

en vue de l'obtention du diplôme de : Philosophiæ Doctor

a été dûment acceptée par le jury d'examen constitué de :

M. PARTOVI NIA Vahid, Doctorat, président

M. LEFEBVRE Mario, Ph. D., membre et directeur de recherche

M. NADEAU Daniel, Ph. D., membre

M. GUILBAULT Jean-Luc, Ph. D., membre externe

DÉDICACE

*À mes parents, mon époux,
et ma grande famille . . .*

REMERCIEMENTS

J'exprime toute ma gratitude, mon respect et ma reconnaissance à mon directeur de recherche, Monsieur le Professeur Mario Lefebvre.

Je remercie Monsieur le Professeur Alain Rousseau, pour son soutien et sa disponibilité durant la période de stage à L'INRS.

Mes remerciements s'adressent également à Messieurs les membres du Jury qui m'ont fait l'honneur de juger ma thèse.

RÉSUMÉ

Les processus de renouvellement filtrés sont utilisés pour prévoir les débits journaliers de cours d'eau un et deux jours à l'avance. Pour ces processus, contrairement au processus de Poisson filtré (Shot noise), le temps entre deux événements consécutifs n'est pas nécessairement exponentiellement distribué, ce qui est plus réaliste. Le modèle est appliqué pour les prévisions des débits des fleuves Delaware et Hudson situés au États-Unis. De meilleurs résultats sont obtenus qu'avec le processus de Poisson filtré qui est souvent utilisé pour modéliser les débits de cours d'eau. Pour obtenir des estimations des probabilités que le débit dépasse un seuil donné à l'instant $t + 1$, étant donné sa valeur à l'instant t , et la distribution du débit au moment de la prochaine augmentation (étant donné que le débit vient juste d'augmenter), deux modèles stochastiques pour les fluctuations des débits sont considérés : le processus de Poisson filtré et le processus de diffusion avec sauts. Les estimations obtenues à partir de la régression linéaire sont aussi considérées à des fins de comparaison. Les paramètres des modèles sont supposés dépendre de la valeur des débits. Les résultats sont appliqués au fleuve Delaware.

Les analyses d'incertitude et de sensibilité permettent de quantifier et d'évaluer l'effet des variations des paramètres d'entrée sur la réponse du modèle. Ces analyses font partie intégrante et nécessaire de la modélisation hydrologique et de la qualité de l'eau. Les méthodes les plus fréquemment utilisées sont : la méthode de moments de premier ordre (MFOSM - *Mean-value first-order second-moment*) et la méthode de moments de second ordre (MSOSM - *Mean-value second-order second-moment*). Ces analyses sont basées sur le calcul d'une fonction de performance approximée par le développement en série de Taylor de premier et second degré au voisinage de la valeur moyenne du jeu de paramètres. L'objectif est de réaliser une analyse de sensibilité et d'incertitude d'un nouveau module spécifique aux milieux humides nouvellement intégré dans HYDROTEL, un modèle hydrologique distribué. Les méthodes de MFOSM et MSOSM sont appliquées sur les débits journaliers simulés par le modèle HYDROTEL à différents segments du bassin versant de la rivière Bécancour, Québec (Canada). Les probabilités de dépassement (EP) d'une valeur du débit donnée sont calculées et comparées en utilisant les simulations de Monte-Carlo (MCS). Les résultats sont analysés par rapport à deux types de milieux humides : isolés et riverains, situés dans trois régions qui divisent le bassin versant de Bécancour. Ces résultats illustrent que les paramètres des milieux humides affectent significativement les variations de débits et que ceux associés aux milieux humides isolés ont un impact plus important par rapport à ceux des milieux humides riverains.

ABSTRACT

Filtered renewal processes are used to forecast daily river flows. For these processes, contrary to filtered Poisson processes, the time between consecutive events is not necessarily exponentially distributed, which is more realistic. The model is applied to obtain one- and two-day-ahead forecasts of the flows of the Delaware and the Hudson Rivers, both located in the United States. Better results are obtained than with filtered Poisson processes, which are often used to model river flows. To obtain estimates of the probability that a river flow will exceed a given threshold at time $t + 1$, given the flow value at time t , and the distribution of the flow at the time of the next increase, given that the flow of a river has just increased, two stochastic models for the fluctuations of the flow are considered: a filtered Poisson process and a diffusion process with jumps. Estimates derived from linear regression are also considered for purposes of comparison. The model parameters are assumed to depend on the flow value. The results are applied to the Delaware River.

Uncertainty and sensitivity analyses provide a framework to quantify and assess the effect of input parameter variations on model response. These analyses are an unavoidable and an integrated part of hydrological and water quality modelling. Mean-value, first-order, second-moment (MFOSM) and mean-value second-order second-moment (MSOSM) methods are frequently used to perform these analyses. They are based on the derivation of a performance function which is approximated using first/second-order Taylor expansion at the mean-value point in the parameter space. The objective is to conduct sensitivity and uncertainty analyses of the wetland modules of HYDROTEL, a continuous, process-based, distributed hydrological model. Following calibration of HYDROTEL on the Bécancour River watershed, Quebec (Canada), the MFOSM and MSOSM methods are applied with respect to simulated daily flows at various river segments. Exceedance Probabilities (EP) of a given streamflow value are computed and compared using Monte Carlo Simulations (MCS). Results are analysed with respect to two types of wetlands, isolated wetlands and riparian wetlands, located in three regions dividing the Bécancour watershed. These results illustrate that the wetland parameters affect significantly streamflows and that those associated with isolated wetlands have a stronger impact when compared to those of riparian wetlands.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vi
TABLE DES MATIÈRES	vii
LISTE DES TABLEAUX	x
LISTE DES FIGURES	xi
LISTE DES SIGLES ET ABRÉVIATIONS	xii
CHAPITRE 1 INTRODUCTION GÉNÉRALE	1
1.1 Éléments de la problématique	1
1.2 Objectifs de recherche	2
1.3 Plan de la thèse	2
I APPLICATION DES PROCESSUS DE RENOUVELLEMENT FIL- TRÉ ET DE DIFFUSION EN HYDROLOGIE	4
CHAPITRE 2 INTRODUCTION	5
CHAPITRE 3 MODÉLISATION DE LA TRAJECTOIRE DU DÉBIT D'UN COURS D'EAU PAR UN PROCESSUS DE RENOUVELLEMENT FILTRÉ	11
3.1 Prévion du débit à l'instant $t + 1$	12
3.2 Prévion du débit à l'instant $t + 2$	13
3.3 Application et résultats	15
3.3.1 Prévion des débits des fleuves Delaware et Hudson	15
CHAPITRE 4 LA PROBABILITÉ THÉORIQUE DE DÉPASSEMENT D'UN SEUIL DE DÉBIT DONNÉ	25
4.1 Application et résultats	28

4.1.1	Implémentation des modèles de distribution de probabilités de dépassement	28
4.1.2	Estimation des paramètres	29
4.1.3	Résultats numériques	31
CHAPITRE 5 LA DISTRIBUTION DU DÉBIT AU MOMENT DE LA PROCHAINE AUGMENTATION		33
5.1	Application et résultats	35
5.1.1	Distribution de $X(\tau_{n+1}) X(\tau_n)$	35
5.1.2	Distribution de Y si $\tau_{n+1} - \tau_n = 1$	39
CHAPITRE 6 CONCLUSION		42
II ANALYSE DE SENSIBILITÉ ET D'INCERTITUDE DES MODULES DE MILIEUX HUMIDES D'UN MODÈLE HYDROLOGIQUE DISTRIBUÉ		45
CHAPITRE 7 INTRODUCTION		46
CHAPITRE 8 L'ANALYSE D'INCERTITUDE ET DE SENSIBILITÉ		48
8.1	Les méthodes existantes de l'analyse d'incertitude	48
8.1.1	La méthode de moments de premier ordre (MFOSM)	50
8.1.2	La méthode de moments de second ordre (MFOSM)	52
8.2	L'analyse de sensibilité	54
8.2.1	L'analyse de sensibilité du premier ordre	55
CHAPITRE 9 ANALYSE D'INCERTITUDE ET DE SENSIBILITÉ DES MODULES DE MILIEUX HUMIDES DU MODÈLE HYDROTEL		57
9.1	Le modèle hydrologique HYDROTEL	57
9.2	Le bassin versant à l'étude	58
9.3	Analyse d'incertitude	58
9.3.1	La méthode de Monte-Carlo	60
9.3.2	La méthode MFOSM	61
9.3.3	La méthode MSOSM	62
9.3.4	Résultats et discussion	63
9.4	Analyse de sensibilité	64
9.4.1	Résultats et discussion	68

CHAPITRE 10 CONCLUSION	70
CHAPITRE 11 CONCLUSION GÉNÉRALE	72
11.1 Synthèse des travaux	72
11.2 Biens livrables	72
RÉFÉRENCES	73

LISTE DES TABLEAUX

Tableau 3.1	Tests d'ajustement de la distribution de la variable aléatoire T dans le cas du fleuve Delaware	17
Tableau 3.2	Critères de performance pour la validation des prévision	19
Tableau 3.3	Critères de comparaison pour les modèles de prévisions dans le cas du fleuve Delaware	20
Tableau 3.4	Tests d'ajustement de la distribution de la variable aléatoire T dans le cas du fleuve Hudson	21
Tableau 3.5	Critères de comparaison pour les modèles de prévisions dans le cas du fleuve Hudson	22
Tableau 4.1	Test d'ajustement pour la distribution des variables aléatoires T_n . .	29
Tableau 4.2	Valeurs numériques des différentes estimations, et fréquences observées correspondantes	32
Tableau 5.1	Test d'ajustement pour la distribution de la variable aléatoire T . . .	36
Tableau 5.2	Second test d'ajustement pour la distribution de la variable aléatoire T	37
Tableau 5.3	Test d'ajustement pour la distribution de la variable aléatoire Y (en milliers de π^3/s)	38
Tableau 5.4	Second test d'ajustement pour la distribution de la variable aléatoire Y (en milliers de π^3/s)	38
Tableau 5.5	Test d'ajustement du khi-deux de la fonction de densité (5.8) de la variable aléatoire $V(10000)$ (en milliers de π^3/s)	39
Tableau 5.6	Test d'ajustement du khi-deux de la fonction de densité (5.12) de la variable aléatoire $V(10000)$ (en milliers de π^3/s)	39
Tableau 5.7	Test d'ajustement du khi-deux de la distribution de la variable aléatoire Y (en milliers de π^3/s), étant donné que $T = 1$	41
Tableau 5.8	Second test d'ajustement du khi-deux de la distribution de la variable aléatoire Y (en milliers de π^3/s), étant donné que $T = 1$	41
Tableau 9.1	Description des paramètres θ_j , leur valeurs de référence θ_{j0} et leurs incertitudes $\Delta\theta_j$ des MH isolés et riverains du bassin versant de la Bécancour	61
Tableau 9.2	Les surfaces des MH drainées et occupées par les secteurs considérés du BV de la rivière de Bécancour	62
Tableau 9.3	Le classement des indices de sensibilité pour chaque sous-tronçon de la rivière Bécancour	69

LISTE DES FIGURES

Figure 2.1	Exemple de la trajectoire du processus de Poisson filtré (de renouvellement filtré) défini par (2.1) et (2.2).	6
Figure 3.1	Les fonctions de répartition empirique et ajustée de la variable aléatoire T dans le cas du fleuve Delaware	17
Figure 3.2	Les fonctions de répartition empirique et ajustée de la variable aléatoire T dans le cas du fleuve Hudson	21
Figure 3.3	Les valeurs observées et prévues du fleuve Delaware à l'instant $t + 1$	23
Figure 3.4	Les valeurs observées et prévues du fleuve Delaware à l'instant $t + 2$	23
Figure 3.5	Les valeurs observées et prévues du fleuve Hudson à l'instant $t + 1$	24
Figure 3.6	Les valeurs observées et prévues du fleuve Hudson à l'instant $t + 2$	24
Figure 5.1	Les fonctions de densité de la v.a $V(10^4) = X(\tau_{n+1}) \{X(\tau_n) = 10^4\}$ déduites de (5.8) (la courbe commençant de 0) et (5.12) avec les paramètres remplacés par leurs estimations ponctuelles.	38
Figure 9.1	Le bassin versant TR1 de la rivière Bécancour et ses sous-bassins ou tronçons TR70, TR102 et TR255	59
Figure 9.2	Comparaison des valeurs des EP sur le TR1 pour la journée du débit maximum résultant des différentes méthodes d'analyse d'incertitude.	64
Figure 9.3	Comparaison des valeurs des EP sur le TR1 pour la journée du débit minimum résultant des différentes méthodes d'analyse d'incertitude.	65
Figure 9.4	Comparaison des valeurs des EP sur le TR70 pour la journée du débit maximum résultant des différentes méthodes d'analyse d'incertitude.	65
Figure 9.5	Comparaison des valeurs des EP sur le TR70 pour la journée du débit minimum résultant des différentes méthodes d'analyse d'incertitude.	66
Figure 9.6	Comparaison des valeurs des EP sur le TR102 pour la journée du débit maximum résultant des différentes méthodes d'analyse d'incertitude.	66
Figure 9.7	Comparaison des valeurs des EP sur le TR102 pour la journée du débit minimum résultant des différentes méthodes d'analyse d'incertitude.	67
Figure 9.8	Comparaison des valeurs des EP sur le TR255 pour la journée du débit maximum résultant des différentes méthodes d'analyse d'incertitude.	67
Figure 9.9	Comparaison des valeurs des EP sur le TR255 pour la journée du débit minimum résultant des différentes méthodes d'analyse d'incertitude.	68

LISTE DES SIGLES ET ABRÉVIATIONS

BV	Bassin Versant
EP	Probabilité de dépassement (Excedence probability)
MCS	Simulation de Monte Carlo (Monte Carlo Simulation)
MFOSM	Méthode des moments d'ordre 1 (Mean-value, first-order, second-moment)
MH	Milieux Humides
MHI	MH isolés
MHR	MH riverains
MSOSM	Méthode des moments d'ordre 2 (Mean-value second-order second-moment)
PPF	Processus de Poisson Filtré
PPR	Processus de Renouvellement Filtré
PDS	Processus de Diffusion avec Sauts
TR	Tronçons
SIG	Système d'Information Géographique
UHRH	Unités Hydrologiques Relativement Homogènes

CHAPITRE 1 INTRODUCTION GÉNÉRALE

1.1 Éléments de la problématique

La modélisation mathématique contribue présentement de façon significative à l'évaluation des changements des caractéristiques hydrologiques des bassins versants. Les caractéristiques des débits et l'effet des milieux humides sur les variations des débits sont des cas particuliers considérés dans cette étude par le biais de la modélisation au moyen des processus stochastiques d'une part, en considérant la prévision à court terme des débits et leur risque de dépassement d'un certain seuil donné, et d'autre part, par l'utilisation d'un métamodèle simulant les débits de bassin versant avec les paramètres considérés comme les seules variables d'entrée.

Les processus de renouvellement filtrés sont considérés, dans un premier temps, comme outils de modélisation et de prévision hydrologique à court terme. Le calcul du risque de crue, par les processus de renouvellement filtrés et de diffusion avec sauts, est considéré dans un second temps. Comparativement aux processus en temps discret tels que les modèles autorégressifs d'ordre un, ce problème est traité au moyen des probabilités de dépassement d'un seuil donné et de la distribution des débits lors de leurs prochaines augmentations.

Par ailleurs, le modèle hydrologique HYDROTEL simulant les débits de bassins versants est considéré avec l'intégration des paramètres spécifiques aux milieux humides. Pour évaluer et quantifier le poids des incertitudes associées aux paramètres sur les débits simulés, l'analyse d'incertitude et l'analyse de sensibilité locale ont été effectuées. Pour des fins de comparaisons de méthodes, l'analyse d'incertitude est réalisée en calculant la probabilité de dépassement d'un seuil de débit donné, les méthodes considérées utilisent un métamodèle issu d'un développement en série de Taylor de premier ou second degré. On considère deux méthodes généralement utilisées, la méthode des moments d'ordre un (MFOSM - *Mean-value first-order second-moment*) et la méthode des moments d'ordre deux (MSOSM - *Mean-value second-order second-moment*); ces deux méthodes seront comparées et validées par les simulations de Monte-Carlo (MCS). L'analyse de sensibilité a pour objectif d'identifier les paramètres les plus sensibles ou influents sur la réponse du modèle.

1.2 Objectifs de recherche

Dans le but de fournir de nouveaux outils robustes aux analystes et instances décisionnelles, pour améliorer les prévisions hydrologiques et gérer les crues et les risques liées aux ressources hydriques, les processus stochastiques caractérisés par des sauts tels les processus de renouvellement filtrés et de diffusion sont proposés comme outil de modélisation des processus hydrologiques.

Les objectifs attendus consistent à : (1) Améliorer significativement les prévisions hydrologiques par l'utilisation du processus de renouvellement filtré, et ce, comparativement aux processus de Poisson filtrés et aux processus autorégressifs d'ordre un. (2) Résoudre le problème du calcul des risques de crues par de nouvelles approches considérées, en particulier, par les processus de Poisson filtrés, les processus de diffusion avec sauts et les processus autorégressifs d'ordre un. Cette étude comparative entre ces différents processus est basée sur une application à des données réelles de débits de cours d'eau.

De plus, nous évaluerons la performance du modèle HYDROTEL en introduisant les paramètres des milieux humides, pour simuler les débits de bassin versant, par une analyse de sensibilité et d'incertitude. Les objectifs étant : (3) De quantifier l'impact de l'incertitude associée aux paramètres d'entrées, sur la réponse du modèle simulant les débits du bassin versant, et ce en calculant les probabilités de dépassement d'un seuil de débit donné, par l'utilisation des méthodes ci-mentionnées. (4) D'identifier les paramètres les plus influents sur la variation des débits.

Finalement, cette thèse se veut originale : (1) de par l'utilisation de nouveaux processus stochastiques pour la modélisation en hydrologie, tels le processus de renouvellement filtré et le processus de diffusion avec sauts qui est souvent utilisé en mathématique financière. (2) de par la proposition d'une nouvelle approche pour estimer la probabilité d'un risque de crue, au moyen des processus stochastiques caractérisés par des sauts. (3) de par l'utilisation de deux méthodes existantes de l'analyse d'incertitude, pour la première fois, sur un nouveau module spécifique aux milieux humides intégré au modèle HYDROTEL pour simuler les débits journaliers d'un bassin versant.

1.3 Plan de la thèse

Cette thèse est composée de deux parties. La première partie (I) est consacrée aux deux premiers objectifs ci-haut mentionnés et la deuxième partie (II) répond aux deux derniers objectifs. On trouve dans la première partie une introduction (chapitre 2) résumant les concepts de base ainsi que les problématiques se rapportant aux trois thèmes considérés dans les cha-

pitres qui suivent : modélisation de la trajectoire du débit d'un cours d'eau par un processus de renouvellement filtré (chapitre 3) ; la probabilité théorique de dépassement d'un seuil de débit donné (chapitre 4) et la distribution du débit au moment de la prochaine augmentation (chapitre 5). La deuxième partie est constituée d'une introduction au sujet (chapitre 7), le contexte théorique de l'analyse de sensibilité et d'incertitude (chapitre 8) et leur application sur les modules humides du modèle HYDROTEL (chapitre 9). Les deux parties se terminent par une conclusion (chapitres 6 et 10 respectivement). Au début et à la fin de la thèse, on trouve une introduction générale et une conclusion générale.

Première partie

APPLICATION DES PROCESSUS DE
RENOUVELLEMENT FILTRÉ ET DE
DIFFUSION EN HYDROLOGIE

CHAPITRE 2 INTRODUCTION

Soit $\{N(t), t \geq 0\}$ un processus de Poisson de taux λ . Le processus de Poisson filtré, appelé aussi shot-noise (voir Parzen (1999) et Lefebvre (2007)), est un processus en temps continu $\{X(t), t \geq 0\}$ défini par

$$X(t) = \sum_{n=1}^{N(t)} w(t, \tau_n, Y_n) \quad (X(t) = 0 \text{ si } N(t) = 0), \quad (2.1)$$

où les variables aléatoires τ_1, τ_2, \dots sont les instants d'arrivée des événements du processus de Poisson et Y_1, Y_2, \dots sont des variables aléatoires supposées indépendantes et identiquement distribuées et qui sont indépendantes du processus $\{N(t), t \geq 0\}$. La fonction $w(., ., .)$ est appelée fonction de réponse. Dans la plupart des applications, la forme de la fonction de réponse est donnée par

$$w(t, \tau_n, Y_n) = Y_n e^{-(t-\tau_n)/c}, \quad (2.2)$$

où c est un paramètre à estimer. La fonction donne la valeur à l'instant t d'un événement de magnitude Y_n du processus de Poisson qui arrive à l'instant τ_n . Les variables aléatoires Y_1, Y_2, \dots sont supposées suivre, en général, une distribution exponentielle de paramètre μ . La trajectoire du processus de Poisson filtré avec cette fonction de réponse est illustrée à la figure (2.1). Cette trajectoire dépend de la forme de la fonction de réponse et non pas de la distribution du temps entre les événements. Par conséquent, la même trajectoire est observée dans le cas où $\{N(t), t \geq 0\}$ est un processus de renouvellement. En effet, le processus de Poisson n'est qu'un cas particulier du processus de renouvellement.

Pour modéliser les débits journaliers de cours d'eau étant donné leurs valeurs observées récentes, les modèles conceptuels et physiques basés sur le processus de Poisson filtré ont été largement utilisés avec succès il y a quelques décennies; voir, par exemple, Weiss (1977), Kelman (1980), Koch (1985) et Konecny (1992). Le processus de Poisson filtré est encore utilisé pour modéliser différents phénomènes en génie civil, par exemple, Yin et al. (2010) et Miyamoto et al. (2012).

En fait, dans les applications hydrologiques en particulier, la forme de la fonction de réponse en (2.2) est présupposée plutôt que d'être justifiée par les observations de la variable d'intérêt. En général, il s'agit tout simplement d'une hypothèse formulée pour obtenir un modèle mathématique simple et traitable. De façon similaire, la distribution exponentielle de la variable Y_n n'est pas vérifiée. Cependant, si on s'intéresse à la distribution de $X(t + \delta)$, basée sur les valeurs du processus jusqu'à l'instant t , cela ne poserait aucun problème puisque

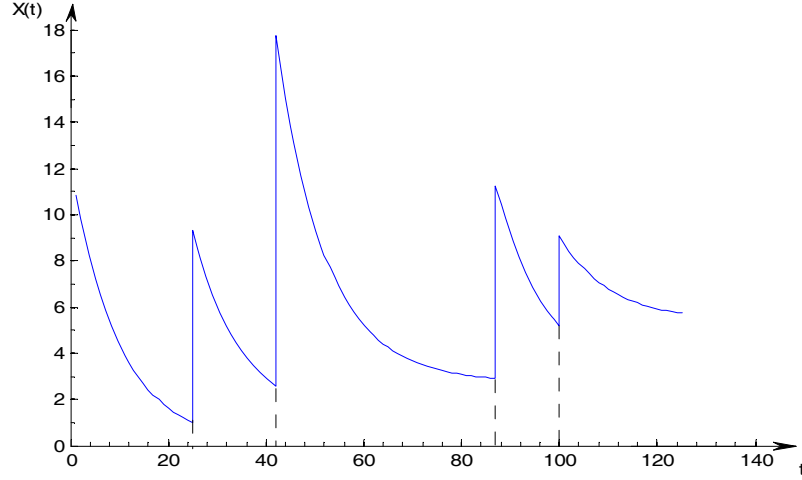


Figure 2.1 Exemple de la trajectoire du processus de Poisson filtré (de renouvellement filtré) défini par (2.1) et (2.2).

la prévision considère seulement la moyenne de Y_n , et sa distribution n'est pas nécessaire. Finalement, l'hypothèse que $\{N(t), t \geq 0\}$ est effectivement un processus de Poisson filtré n'est pas testée. Ce n'est qu'une autre hypothèse simplificatrice, parce qu'on peut utiliser par la suite les propriétés intéressantes du processus de Poisson, notamment la propriété des accroissements indépendants et stationnaires.

Avec la fonction de réponse en (2.2), l'effet sur la valeur du processus d'un événement qui arrive à l'instant τ_n est maximale à cet instant et $\{X(t), t \geq 0\}$ a des discontinuités aux instants d'arrivée des événements du processus de Poisson filtré. Dans le cas des applications en hydrologie qui consistent à modéliser les débits de cours d'eau, l'observation d'hydrogrammes réels montre qu'il y a une plus ou moins longue période durant laquelle le débit augmente du minimum au maximum. Même si les données sur les débits sont souvent disponibles sur une base quotidienne, on ne peut pas observer les augmentations très soudaines du débit, suivies par une décroissance exponentielle. En fait, le débit prend en moyenne deux à trois jours pour atteindre la valeur maximale et décroît ensuite plus ou moins rapidement (Lefebvre and Bensalma (2015)).

Pour obtenir un modèle plus réaliste pour les variations des débits de cours d'eau, en tenant compte de la période de l'augmentation du débit, Lefebvre and Guilbault (2008) ont utilisé la fonction de réponse suivante :

$$w(t, \tau_n, Y_n) = Y_n(t - \tau_n)^d e^{-(t - \tau_n)/c}, \quad (2.3)$$

où d est un paramètre positif. Avec cette fonction de réponse, le débit de cours d'eau augmente immédiatement après l'instant τ_n , mais le maximum est atteint après dc unités de temps (en jours, généralement), et après il commence à décroître. Leur travail a été amélioré par Lefebvre and Bensalma (2015), qui ont fourni une méthode pour estimer le paramètre d . Les auteurs ont trouvé, pour l'application qu'ils ont considérée, que la valeur de d appartient à l'intervalle $(0, 1)$.

Le processus de Poisson étant une chaîne de Markov en temps continu, le temps que le processus passe dans un état donné est d'une distribution exponentielle, avec une fonction de densité strictement décroissante. Soit T_1, T_2, \dots les temps entre deux événements successifs (ou les temps d'inter-occurrence des événements). Dans le cas des débits de cours d'eau, le temps d'inter-occurrence est généralement une variable aléatoire avec une fonction de densité qui augmente en premier. Par conséquent, l'hypothèse que $\{N(t), t \geq 0\}$ est un processus de Poisson est pour le moins douteuse.

Lefebvre (2003) teste l'hypothèse que les variables T_n sont distribuées exponentiellement. Il trouve, pour le fleuve Delaware, que l'hypothèse n'est pas vérifiée. Il montre que les distributions gamma et de Rayleigh ajustent adéquatement les données. L'objectif était de prévoir les débits de pointe du fleuve Delaware, qui est un problème difficile dû à la faible corrélation entre les pointes successives. Pour prévoir les débits de pointe, l'auteur a utilisé le processus de renouvellement filtré comme modèle pour les débits de cours d'eau. L'avantage d'utiliser la distribution de Rayleigh pour le temps d'inter-occurrence des événements est que le processus de renouvellement filtré peut être transformé en processus de Poisson filtré en travaillant sur une échelle de temps différente.

Les processus de Poisson et de renouvellement ont été utilisés en hydrologie pour modéliser les faibles débits (voir par exemple, Loaiciga and Leipnik (1996), Yagouti et al. (2001)) et dans divers phénomènes physiques comme le bruit du trafic routier (voir Marcus (1973)). Andersen (1979, 1981) a trouvé de nouveaux résultats théoriques pour le processus de renouvellement filtré (voir aussi les références dans les deux travaux).

Le premier objectif dans ce présent travail est de montrer qu'on peut obtenir des prévisions à court terme des débits de cours d'eau plus précises par le processus de renouvellement filtré que par le processus de Poisson filtré. Pour cela, on développe d'abord une formule pour prévoir le débit à l'instant $X(t + \delta)$ étant donné l'historique du processus jusqu'à l'instant t . Du fait qu'on ne peut pas utiliser la propriété du non-vieillessement de la distribution exponentielle, le calcul de l'estimateur de $X(t + \delta)$ est plus difficile. Les valeurs du débit avant l'instant t doivent être prises en considération, à l'opposé du cas de processus de Poisson filtré, pour lequel les prévisions dépendent seulement de $X(t)$.

Par ailleurs, dans le but de voir clairement l'amélioration obtenue par l'utilisation d'une distribution plus réaliste du temps d'inter-occurrence des événements, on considérera la fonction de réponse définie en (2.2). Comme mentionné précédemment, le processus de Poisson filtré avec cette fonction de réponse est souvent utilisé en hydrologie et les formules de prévision sont faciles à calculer et implémenter. Alors, il sera facile de juger de la qualité du modèle qu'on propose.

On montre (voir Lefebvre and Bensalma (2014, doi :10.1155/2014/593243)) que dans le cas des fleuves Delaware et Hudson, la distribution exponentielle n'est pas un bon modèle pour les variables T_n , et ce, en considérant les valeurs des débits durant une année de calendrier. Cependant, en pratique, les paramètres du modèle dépendent de la valeur courante du débit. On montrera plus loin dans cette étude que la distribution exponentielle s'ajuste bien aux données lorsque la valeur de $X(t)$ est grande. Alors on peut considérer le processus de Poisson filtré simple pour estimer la probabilité que le débit à l'instant $t + 1$ dépassera un certain seuil donné, en connaissant la valeur de $X(t)$ et si $X(t)$ est suffisamment grand.

Une autre classe de processus stochastiques qui a été largement utilisée dans les mathématiques financières pour modéliser les prix des actions est celle des processus de diffusion avec sauts. Cont and Tankov (2004) fournissent presque 400 références sur le sujet qui est basé sur l'article de Merton (1976).

Soit $\{D(t), t \geq 0\}$ un processus de diffusion et $\{N(t), t \geq 0\}$ un processus de Poisson. Le processus de diffusion avec sauts $\{X_D(t), t \geq 0\}$ est défini par

$$X_D(t) = D(t) + \sum_{n=1}^{N(t)} Y_n \quad \text{pour } t \geq 0, \quad (2.4)$$

où les variables Y_n sont définies précédemment. On ajoute donc un processus de Poisson composé (qui est un processus de Poisson filtré particulier) au processus de diffusion. Les deux processus stochastiques sont supposés indépendants. Contrairement au processus de Poisson filtré défini par (2.1) avec la fonction de réponse (2.2) :

$$X(t) = \sum_{n=1}^{N(t)} Y_n e^{-(t-\tau_n)/c} \quad (X(t) = 0 \text{ si } N(t) = 0), \quad (2.5)$$

qui se comporte de façon déterministe entre les événements, la trajectoire du processus de diffusion avec sauts est complètement aléatoire. On supposera que $\{D(t), t \geq 0\}$ est un processus de Wiener avec une dérive négative. Par conséquent, après l'arrivée d'un événement, $X_D(t)$ tend probablement à décroître. Cependant, $X_D(t)$ peut théoriquement continuer à augmenter pendant un certain temps. Par ailleurs, la décroissance ne se fait pas toujours

exponentiellement, comme dans le cas du processus de Poisson filtré (5.1). On peut donc affirmer que le processus de diffusion avec sauts peut être un modèle plus réaliste que le processus défini en (5.1) pour les variations des débits de cours d'eau.

Le deuxième objectif est d'estimer la probabilité que le débit augmente d'au moins y unités en un jour, lorsque le débit courant est élevé, tel qu'il y ait un sérieux risque de crue, et ce, en utilisant les deux modèles (5.1) et (5.10). Avec des données observées sur une longue période, on peut estimer les paramètres du modèle ainsi que la probabilité de dépassement dépendamment de différentes valeurs de y . Ces estimations seront comparées avec les fréquences observées durant la même période dans le cas du fleuve Delaware.

On s'intéresse, pour le troisième objectif, à la distribution de probabilité de la variable aléatoire $X(\tau_{n+1})$, étant donnée la valeur de $X(\tau_n)$. On cherche à déterminer la distribution du débit au moment du prochain signal, étant donné que le débit vient juste d'augmenter à $X(\tau_n)$. Dans le cas du processus de Poisson filtré en (5.1), la forme de la fonction de réponse implique que le débit diminue de l'instant t à $t + 1$ s'il n'y a pas d'événement dans cet intervalle de temps $(t, t + 1]$. Si $X(t + 1) \geq X(t)$, on peut conclure qu'il y a eu au moins un événement entre t et $t + 1$. Théoriquement, le débit peut diminuer même s'il y a un signal dans l'intervalle $(t, t + 1]$, si le signal en question est faible. On néglige ici cette possibilité. Puisque les valeurs de débits sont généralement mesurées sur une base journalière, on ne peut pas déterminer le nombre d'événements qui arrivent dans l'intervalle $(t, t + 1]$. Donc, chaque augmentation de la valeur du débit signifie l'arrivée d'un événement entre t et $t + 1$. Lorsqu'on utilise le processus de diffusion avec sauts pour modéliser les fluctuations des débits de cours d'eau, le débit peut être plus élevé à l'instant $t + 1$ qu'à l'instant t , même s'il y a une absence d'événements dans l'intervalle $(t, t + 1]$. Cependant, en supposant un paramètre de dérive négatif du processus de Wiener, la probabilité d'une telle augmentation dans le cas où il n'y a pas de signaux entre t et $t + 1$ est négligeable. De là, on suppose qu'un signal se produit entre t et $t + 1$ si et seulement si $X(t + 1) \geq X(t)$, comme dans le cas du processus de Poisson filtré. Une application est présentée dans le cas du fleuve Delaware pour estimer la distribution de $X(\tau_{n+1}) | X(\tau_n)$ ainsi que la distribution des variables aléatoires Y_n sous l'hypothèse que $\tau_{n+1} - \tau_n = 1$.

Les objectifs énoncés dans cette première partie concernant la modélisation hydrologique au moyen des processus stochastiques se rapportent aux trois chapitres 3, 4 et 5, et constituent trois livrables sous forme d'articles scientifiques de revues :

Bien livrable 1 (Chapitre 3)

Utilisation du processus de renouvellement filtré (PRF) comme modèle de prévision des

débits de cours d'eau. L'objectif est de proposer quelques issues pour la solution aux problèmes de gestion de l'eau, que ce soit pour la prévision des apports aux réservoirs afin de planifier la production hydroélectrique, ou pour l'opération des barrages publics afin de gérer les crues, qui sont des pratiques courantes et surtout incontournables en hydrologie.

Biens livrables 2 et 3 (Chapitre 4 et 5)

Utilisation du processus de renouvellement filtré (PRF) et du processus de diffusion avec sauts (PDS) pour le calcul des risques d'inondation. Pour ce faire, on estime les probabilités de débits de crues et la distribution du débit lors de sa prochaine augmentation, ce qui relève d'une grande importance pour les ingénieurs en hydrologie. En particulier, pour prévoir les dangers impliquant des vies humaines en cas de crues et aussi de l'intérêt économique relatif à l'exploitation de certains ouvrages nécessaires à la prévention des événements extrêmes.

Enfin, le chapitre 6 présente les principales conclusions de cette première partie de la thèse.

CHAPITRE 3 MODÉLISATION DE LA TRAJECTOIRE DU DÉBIT D'UN COURS D'EAU PAR UN PROCESSUS DE RENOUVELLEMENT FILTRÉ

Soit $X(t)$ le débit d'un certain cours d'eau à l'instant t . On suppose que

$$X(t) = \sum_{n=1}^{N(t)} Y_n e^{-(t-\tau_n)/c} \quad (X(t) = 0 \text{ si } N(t) = 0), \quad (3.1)$$

tel que $\{N(t), t \geq 0\}$ est un processus de renouvellement. Les τ_n sont les instants d'arrivées des événements du processus $\{N(t), t \geq 0\}$ et les temps $T_n = \tau_n - \tau_{n-1}$ entre les renouvellement successifs sont des variables aléatoires indépendantes et identiquement distribuées. Finalement, les variables aléatoires (positives) Y_1, Y_2, \dots sont indépendantes et identiquement distribuées et indépendantes du processus $\{N(t), t \geq 0\}$.

Ce processus est donc caractérisé par des accroissements soudains du débit causés par les événements qui arrivent aux instants τ_1, τ_2, \dots . Le cycle de renouvellement est défini comme étant le temps T_n entre deux pointes (valeurs maximales du débit) consécutives. Pour le temps d'inter-arrivée, les distributions suivantes (en plus de la distribution exponentielle) seront considérées : lognormale, gamma, Weibull et Rayleigh.

Pour pouvoir prévoir le débit à l'instant $t + \delta$, on doit d'abord estimer le paramètre c qui apparaît dans la fonction de réponse. S'il n'y a pas de signaux entre t et $t + 1$, alors $X(t + 1)$ est donnée par

$$X(t + 1) = e^{-1/c} X(t). \quad (3.2)$$

De là, pour estimer le paramètre c , on peut considérer tous les jours où le débit décroît dans la série de données et calculer le quotient $R = X(t + 1)/X(t)$. L'estimation de c est donc calculée à partir de la moyenne arithmétique \bar{R} de ce quotient :

$$\hat{c} = -\frac{1}{\ln \bar{R}}. \quad (3.3)$$

Dans le cas des variables aléatoires Y_n , qui constituent un échantillon aléatoire de la variable aléatoire Y , leur distribution exacte n'est pas nécessaire. En effet, seulement la valeur espérée de Y sera utilisée pour prévoir la valeur de $X(t + \delta)$. Par conséquent, il suffit tout simplement de calculer l'augmentation moyenne du débit durant la période de calibration du modèle.

On obtient dans ce qui suit des formules pour estimer le débit un et deux jours à l'avance. Ces formules peuvent être utilisées avec n'importe quelle distribution de T , correspondant

aux variables aléatoires T_n .

3.1 Prévision du débit à l'instant $t + 1$

On essaie d'abord d'estimer le débit à l'instant $t+1$, connaissant toute l'information disponible jusqu'à t . Pour ce faire, on suppose qu'il y aura au plus un événement dans l'intervalle $(t, t + 1]$. En tout cas, comme mentionné ci-dessus, en pratique, les valeurs de débit sont en général disponibles sur une base quotidienne. Par conséquent, il n'est pas vraiment possible de déterminer s'il y a plus d'un événement qui s'est produit dans l'intervalle $(t, t + 1]$. Lorsque cela arrive, ces événements peuvent être considérés comme un seul événement important.

Mis à part le cas où le temps d'inter-arrivée T a une distribution exponentielle, on doit prendre en compte l'historique du processus pour calculer la probabilité qu'il y aura bien un événement dans l'intervalle $(t, t + 1]$. Soit

$$U_i = \tau_{N(t)+i} - \tau_{N(t)+i-1} \quad \text{pour } i = 1, 2, \dots \quad (3.4)$$

U_i désigne le temps écoulé entre le $(i - 1)^{\text{e}}$ et le i^{e} événement après le plus récent qui a eu lieu avant ou à l'instant t . Les variables aléatoires T et U_1 sont identiquement distribuées.

On doit calculer

$$p_k := P[U_1 \leq k + 1 \mid U_1 > k], \quad (3.5)$$

où k est le nombre (entier) de jours depuis l'observation du plus récent signal. On a

$$p_k = \frac{\int_k^{k+1} f_{U_1}(u) du}{\int_k^{\infty} f_{U_1}(u) du}. \quad (3.6)$$

Par l'utilisation d'un logiciel mathématique, on peut obtenir des valeurs numériques de cette probabilité pour toute fonction de densité f_{U_1} et toute valeur de k .

Lorsque la variable T est exponentiellement distribuée, il est bien connu que, sachant qu'il y a un signal (un événement) entre t et $t + 1$, la densité conditionnelle de T est uniforme sur l'intervalle $(t, t + 1]$, tel que le signal en question arrive en moyenne à l'instant $t + \frac{1}{2}$ (Lefebvre (2007)). Dans le cas général, on doit calculer la fonction de densité conditionnelle de U_1 dans l'intervalle $(k, k + 1]$, étant donné que $U_1 > k$.

S'il y a effectivement un signal entre t et $t + 1$, la valeur prévue Z_k du débit à l'instant $t + 1$ serait

$$Z_k := X(t) e^{-1/c} + E[Y] \frac{1}{P[U_1 > k]} \int_k^{k+1} e^{-(k+1-u)/c} f_{U_1}(u) du. \quad (3.7)$$

Par conséquent, la prévision du débit à l'instant $t + 1$, compte tenu de tout l'historique du processus jusqu'à l'instant t , est donnée par

$$\begin{aligned}\widehat{X}(t+1) &= X(t) e^{-1/\hat{c}} (1 - p_k) + Z_k p_k \\ &= X(t) e^{-1/\hat{c}} + p_k E[Y] \frac{1}{P[U_1 > k]} e^{-(k+1)/\hat{c}} \int_k^{k+1} e^{u/\hat{c}} f_{U_1}(u) du.\end{aligned}\quad (3.8)$$

3.2 Prévision du débit à l'instant $t + 2$

Soit $U_{1,k} = U_1 \mid \{U_1 > k\}$. Pour prévoir le débit à l'instant $t + 2$, en considérant toute l'information disponible jusqu'à l'instant t , on a besoin de la distribution de la variable aléatoire $S_k := U_{1,k} + U_2$, où U_2 est défini en (3.4).

On suppose ici que la probabilité qu'il y aura plus de deux signaux dans l'intervalle $(t, t + 2]$ est négligeable (indépendamment de la valeur de k). La fonction de densité de S_k est donnée par la convolution des fonctions de densité de $U_{1,k}$ et U_2 :

$$f_{S_k}(s + k) = \int_k^\infty f_{U_{1,k}}(u) f_{U_2}(s + k - u) du. \quad (3.9)$$

Puisque U_2 est une variable aléatoire non négative, on peut écrire que

$$f_{S_k}(s + k) = \int_k^{s+k} f_{U_{1,k}}(u) f_{U_2}(s + k - u) du \quad \text{pour } s > 0. \quad (3.10)$$

Dans le cas où T a une distribution exponentielle de paramètre λ , on sait que $(S_k - k)$ a une distribution gamma de paramètres $\alpha = 2$ et λ . Dans les autres cas, on doit généralement évaluer l'intégrale précédente numériquement.

La valeur de $X(t + 2)$ est donnée par

$$X(t+2) = \begin{cases} e^{-2/c} X(t) & \text{si } U_{1,k} > k + 2, \\ e^{-2/c} X(t) + Y_1^* e^{-(k+2-U_{1,k})/c} & \text{si } U_{1,k} \leq k + 2 \\ & \text{et } U_2 > 2 - (U_{1,k} - k), \\ e^{-2/c} X(t) + Y_1^* e^{-(k+2-U_{1,k})/c} + Y_2^* e^{-(2-U_2)/c} & \text{si } U_{1,k} + U_2 \leq k + 2 \\ & \text{(et } U_3 > 2 - (U_{1,k} - k) - U_2), \end{cases} \quad (3.11)$$

où Y_i^* représente la taille (ou l'amplitude) du i^e signal dans l'intervalle $(t, t + 2]$.

Tel que mentionné ci-dessus, on suppose que

$$P[U_3 > 2 - (U_{1,k} - k) - U_2] \simeq 1. \quad (3.12)$$

On doit d'abord calculer

$$p_{1,k} := P[U_{1,k} \in (k, k + 2]]. \quad (3.13)$$

Lorsque U_1 suit une distribution exponentielle de paramètre λ , on a

$$p_{1,k} = P[U_1 \in (0, 2]] = 1 - e^{-2\lambda}. \quad (3.14)$$

En général, on a

$$p_{1,k} = \frac{\int_k^{k+2} f_{U_1}(u) du}{\int_k^\infty f_{U_1}(u) du}. \quad (3.15)$$

$$\begin{aligned} p_{2,k} &:= P[U_{1,k} + U_2 \in (k, k + 2]] \equiv P[S_k \in (k, k + 2]] \\ &= \int_0^2 f_{S_k}(s + k) ds \\ &= \int_0^2 \left[\int_k^{s+k} f_{U_{1,k}}(u) f_{U_2}(s + k - u) du \right] ds. \end{aligned} \quad (3.16)$$

On peut également, en utilisant le fait que $U_{1,k}$ et U_2 (≥ 0) sont des variables aléatoires indépendantes, écrire que

$$\begin{aligned} p_{2,k} &:= P[U_{1,k} + U_2 \in (k, k + 2]] \\ &= \int_k^{k+2} P[U_{1,k} + U_2 \in (k, k + 2] \mid U_{1,k} = u] f_{U_{1,k}}(u) du \\ &= \int_k^{k+2} P[U_2 \in (k - u, k + 2 - u)] f_{U_{1,k}}(u) du \\ &= \int_k^{k+2} \left[\int_0^{k+2-u} f_{U_2}(v) dv \right] f_{U_{1,k}}(u) du. \end{aligned} \quad (3.17)$$

Finalement, par indépendance,

$$\begin{aligned}
P[\{U_{1,k} \leq k+2\} \cap \{U_2 > 2 - (U_{1,k} - k)\}] &= \int_k^{k+2} P[U_2 \in (k+2-u, \infty)] f_{U_{1,k}}(u) du \\
&= \int_k^{k+2} \left[\int_{k+2-u}^{\infty} f_{U_2}(v) dv \right] f_{U_{1,k}}(u) du \\
&= \int_k^{k+2} \left[1 - \int_0^{k+2-u} f_{U_2}(v) dv \right] f_{U_{1,k}}(u) du \\
&= p_{1,k} - p_{2,k}.
\end{aligned} \tag{3.18}$$

La prévision du débit à l'instant $t+2$, connaissant tout l'historique du processus jusqu'à l'instant t , est donc donnée par

$$\begin{aligned}
\widehat{X}(t+2) &= X(t) e^{-2/\hat{c}} (1 - p_{1,k}) \\
&\quad + \left\{ X(t) e^{-2/\hat{c}} + E[Y] \int_k^{k+2} e^{-(k+2-u)/\hat{c}} f_{U_{1,k}}(u) du \right\} (p_{1,k} - p_{2,k}) \\
&\quad + \left\{ X(t) e^{-2/\hat{c}} + E[Y] \int_k^{k+2} e^{-(k+2-u)/\hat{c}} f_{U_{1,k}}(u) du \right. \\
&\quad \quad \left. + E[Y] \int_k^{k+2} \left[\int_0^{k+2-u} e^{-(2-v)/\hat{c}} f_{U_2}(v) dv \right] f_{U_{1,k}}(u) du \right\} p_{2,k} \\
&= X(t) e^{-2/\hat{c}} + \left\{ E[Y] \int_k^{k+2} e^{-(k+2-u)/\hat{c}} f_{U_{1,k}}(u) du \right\} p_{1,k} \\
&\quad + \left\{ E[Y] \int_k^{k+2} \left[\int_0^{k+2-u} e^{-(2-v)/\hat{c}} f_{U_2}(v) dv \right] f_{U_{1,k}}(u) du \right\} p_{2,k} \\
&= X(t) e^{-2/\hat{c}} + \left\{ E[Y] e^{-(k+2)/\hat{c}} \int_k^{k+2} e^{u/\hat{c}} f_{U_{1,k}}(u) du \right\} p_{1,k} \\
&\quad + \left\{ E[Y] \int_k^{k+2} \left[e^{-2/\hat{c}} \int_0^{k+2-u} e^{v/\hat{c}} f_{U_2}(v) dv \right] f_{U_{1,k}}(u) du \right\} p_{2,k}.
\end{aligned} \tag{3.19}$$

3.3 Application et résultats

3.3.1 Prévision des débits des fleuves Delaware et Hudson

Pour évaluer la qualité des prévisions obtenues avec le processus de renouvellement filtré, on doit appliquer les formules développées dans ce chapitre (3) (bien livrable 1) à des données réelles. Comme dans le travail de Lefebvre and Guilbault (2008), on choisit les fleuves Delaware et Hudson, situées aux États-Unis, aux stations de jaugeage (01438500) à Montague et (01315500) à North Creek respectivement. Les valeurs des débits sont disponibles sur Internet à l'adresse suivante : <http://nwis.waterdata.usgs.gov>. Pour calibrer le modèle, on a utilisé

les données d'octobre 2008 à septembre 2009.

Remarques :

(i) Quand on applique la théorie sur des données réelles, les choses ne sont pas aussi simples que les modèles mathématiques le suggèrent. En regardant l'hydrogramme réel du fleuve Delaware, on observe un certain nombre de faibles pointes qui se produisent lorsque le débit diminue et pour lesquelles la période d'augmentation dure seulement une journée avant que le débit reprenne son déclin. De même, il y a parfois des petites augmentations de la valeur du débit survenant juste après qu'un minimum est observé et qui ne durent qu'une seule journée. Dans la construction de l'ensemble de données, il y a toujours une part de subjectivité. On a décidé de négliger ces faibles pointes, considérant ainsi uniquement les pointes qui apparaissent clairement dans l'hydrogramme.

(ii) Puisque le débit est observé sur une base quotidienne, on a discrétisé l'ensemble des valeurs possibles prises par le temps d'inter-arrivées T . On a calculé le nombre k de jours écoulés entre les pointes consécutives de l'ensemble de données. Dans le cas de la première pointe observée, k est le nombre de jours écoulés depuis le début de la période de calibrage. Ensuite, k est obtenu en soustrayant les temps d'arrivée des pointes consécutives. Notons que k est supérieur ou égal à 1.

Dans le cas du fleuve Delaware, on trouve que la valeur moyenne \bar{t} du temps d'inter-arrivées dans la série des données est égale à 6,8889 jours et l'écart-type des observations est donné par $s_T = 3,8977$ jours. En se basant sur ces valeurs, on peut conclure que T ne suit probablement pas une distribution exponentielle. En effet, rappelons que la moyenne et l'écart-type d'une loi exponentielle de paramètre λ égalent tous les deux $1/\lambda$.

La valeur de l'estimation ponctuelle du paramètre c (voir (4.17)) est $\hat{c} = 9,2592$, et la valeur moyenne de l'amplitude du signal Y des signaux est donnée par $\bar{y} = 984,6121$ pieds cubes par seconde.

Pour la distribution de la variable aléatoire T , on teste les modèles suivants :

- la distribution exponentielle de paramètre λ ,
- la distribution lognormale de paramètres μ et σ ,
- la distribution gamma de paramètre de forme α et paramètre d'échelle λ ,
- la distribution Weibull de paramètre de forme λ et paramètre d'échelle κ ,
- la distribution Rayleigh de paramètre σ .

Le tableau (3.1) résume les résultats du test d'ajustement du khi-deux pour chaque distribution et donne les estimations des paramètres respectifs. On trouve, comme attendu, que la distribution exponentielle est rejetée, alors que les autres distributions sont toutes acceptées,

selon les valeurs p des tests. Les courbes de la fonction de répartition cumulative empirique ainsi que les fonctions de répartition des différents modèles considérés ci-dessus sont présentées dans la figure (3.1). On voit clairement que la distribution exponentielle ne s'ajuste pas aux données aussi bien que les quatre autres modèles.

Tableau 3.1 Tests d'ajustement de la distribution de la variable aléatoire T dans le cas du fleuve Delaware

Distributions	Exponentielle	Lognormale	Gamma	Weibull	Rayleigh
Paramètres	1/6,8889	1,7574 ; 0,6187	3,0544 ; 2,2554	7,7974 ; 1,8975	5,5817
Valeurs p	0,0344	0,7887	0,7723	0,6927	0,7008

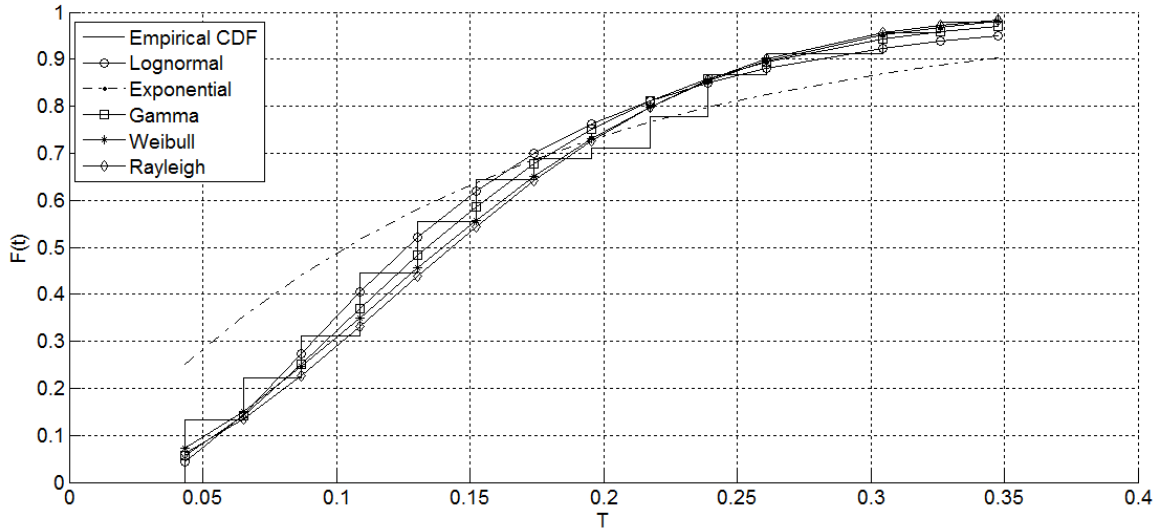


Figure 3.1 Les fonctions de répartition empirique et ajustée de la variable aléatoire T dans le cas du fleuve Delaware

Dans le cas du processus de Poisson filtré (5.1), pour estimer le débit à l'instant $t+1$, on calcule l'espérance de $X(t+1)$ sachant l'historique des observations. En raison de la propriété de non-vieillessement de la distribution exponentielle, on trouve que cette espérance conditionnelle dépend seulement de la valeur la plus récente du débit $X(t)$. On obtient (voir Lefebvre and Guibault (2008)) que

$$E[X(t+1) | X(t)] = e^{-1/c} X(t) + E[X(1)], \quad (3.20)$$

où

$$E[X(1)] = \frac{\lambda c}{\mu} (1 - e^{-1/c}). \quad (3.21)$$

De façon similaire, pour estimer le débit à l'instant $t + 2$, on calcule

$$\begin{aligned} E[X(t+2) | X(t)] &= E \left[\sum_{n: 0 \leq \tau_n \leq t+2} Y_n e^{-(t+2-\tau_n)/c} \mid X(t) \right] \\ &= e^{-2/c} E \left[\sum_{n: 0 \leq \tau_n \leq t} Y_n e^{-(t-\tau_n)/c} \mid X(t) \right] \\ &\quad + E \left[\sum_{n: t < \tau_n \leq t+2} Y_n e^{-(t+2-\tau_n)/c} \mid X(t) \right] \\ &= e^{-2/c} X(t) + E[X(2)], \end{aligned} \quad (3.22)$$

où

$$E[X(2)] = \frac{\lambda c}{\mu} (1 - e^{-2/c}). \quad (3.23)$$

Par ailleurs, on a que

$$\lim_{t \rightarrow \infty} E[X(t)] = \frac{\lambda c}{\mu}, \quad (3.24)$$

$$\lim_{t \rightarrow \infty} \text{Var}[X(t)] = \frac{\lambda c}{\mu^2} \quad (3.25)$$

et

$$\lim_{t \rightarrow \infty} \rho_{X(t), X(t+\delta)} = e^{-\delta/c}, \quad (3.26)$$

où $\rho_{X(t), X(t+\delta)}$ dénote le coefficient de corrélation de $X(t)$ et $X(t + \delta)$.

En utilisant ces trois équations, on peut estimer le paramètre λ du processus de Poisson, le paramètre μ de la distribution exponentielle supposée pour Y_n , et le paramètre c dans la fonction de réponse. En fait, on n'a pas besoin des estimations de λ et μ pour prévoir le débit. À partir de (3.24), on déduit qu'on peut simplement estimer le quotient $\lambda c / \mu$ qui apparaît dans (3.21) et (3.23) par la valeur moyenne des données observées lorsque le processus est en état stationnaire. Finalement, pour estimer le paramètre c , on calcule d'abord le coefficient de corrélation empirique r_1 entre les valeurs des débits aux instants t et $t + 1$, et on pose $\hat{c} = -1 / \ln(r_1)$ (voir (3.26)).

Remarque : Si on s'intéresse à la prévision du débit à l'instant $t + \delta$, on peut estimer c en calculant la valeur empirique de $\rho_{X(t), X(t+\delta)}$ et en utilisant (3.26).

On peut maintenant comparer les prévisions calculées à partir du processus de Poisson filtré

et le processus de renouvellement filtré avec les différentes distributions considérées pour le temps d'inter-arrivée T . On utilise les formules obtenues précédemment pour prévoir le débit du fleuve Delaware sur une période de 89 jours de février à avril 2010.

Pour évaluer la qualité des prévisions issues des deux modèles, on considère quatre critères de performance fréquemment utilisés en hydrologie : l'*erreur moyenne absolue en pourcentage* (Mape), le *critère de Nash* (Nash), le *critère de pointe* (Pc) et le *coefficient de corrélation* (Corr) entre les valeurs observées et prévues. Les formulations mathématiques de ces critères sont présentées au tableau (3.2) en annexe A.

Tableau 3.2 Critères de performance pour la validation des prévision

	Critère	Formulation mathématique	intervalle de valeur	valeur pour une prévision parfaite
Mape	Erreur moyenne absolue en %	$\frac{1}{n} \sum_{i=1}^n \left \frac{\hat{X}_i(t) - X_i(t)}{\hat{X}_i(t)} \right \times 100$	$[0, \infty[$	0
Nash	Critère de Nash	$1 - \frac{\sum_{i=1}^n (\hat{X}_i(t) - X_i(t))^2}{\sum_{i=1}^n (X_i(t) - \bar{X}_i(t))^2}$	$] - \infty, 1]$	1
Pc	Critère de pointe*	$1 - \frac{\left(\sum_{i=1}^m (\hat{X}_i(t) - X_i(t))^2 X_i(t)^2 \right)^{1/4}}{\sqrt{\sum_{i=1}^m X_i(t)^2}}$	$[0, \infty[$	0
Corr	Coefficient de corrélation	$\frac{\sum_{i=1}^n (\hat{X}_i(t) - \bar{\hat{X}}_i(t)) \sum_{i=1}^n (X_i(t) - \bar{X}_i(t))}{\sqrt{\sum_{i=1}^n (\hat{X}_i(t) - \bar{\hat{X}}_i(t))^2 \sum_{i=1}^n (X_i(t) - \bar{X}_i(t))^2}}$	$[-1, 1]$	1

* m est le nombre de données qui ont une valeur supérieure au tiers de la valeur moyenne des débit durant la crue.

Le critère de Nash est basé sur l'erreur quadratique moyenne et est utilisé pour évaluer le pouvoir prédictif des modèles hydrologiques. Il évalue la qualité des prévisions à partir des différences entre les valeurs quotidiennes attendues et la moyenne des valeurs observées. Il est égal à 1 dans le cas d'un ajustement parfait. Pour sa part, le critère de pic (ou de pointe) est utilisé pour mesurer la qualité des prévisions au cours de la période de pointe critique. Lorsque sa valeur est plus proche de 0, les prévisions seront meilleures.

Le tableau (3.3) présente les résultats obtenus avec les deux modèles, le processus de Poisson filtré (PPF) et le processus de renouvellement filtré (PRF). On remarque que les critères Mape, Nash et Corr donnent pratiquement les mêmes valeurs pour les deux modèles. Cependant, le PRF est mieux que le PPF lorsqu'on considère le critère de pointe (Pc). Ce critère est particulièrement important parce que des prévisions précises sont vraiment nécessaires

au cours de la période de pointe.

Tableau 3.3 Critères de comparaison pour les modèles de prévisions dans le cas du fleuve Delaware

		Processus de renouvellement filtré				Processus de Poisson filtré
		Lognormale	Gamma	Weibull	Rayleigh	Exponentielle
$X(t + 1)$	Mape	12,58%	12,49%	12,55%	12,56%	12,76%
	Nash	0,8288	0,8290	0,8287	0,8287	0,8356
	Pc	0,1674	0,1664	0,1663	0,1669	0,2522
	Corr	0,9188	0,9186	0,9180	0,9188	0,9187
$X(t + 2)$	Mape	20,36%	20,33%	20,34%	20,34%	21,83%
	Nash	0,5447	0,5445	0,5445	0,5445	0,5573
	Pc	0,2930	0,2893	0,2869	0,2863	0,3344
	Corr	0,7735	0,7732	0,7731	0,7730	0,7734

En utilisant la même période de prévision que dans le cas du fleuve Delaware, on applique maintenant les formules de prévisions des deux modèles PPF et PRF sur le fleuve Hudson. La valeur estimée de c est $\hat{c} = 9,1172$ et la valeur moyenne de l'amplitude des signaux Y est $\bar{y} = 229,2533$ pieds cubes par seconde, qui est beaucoup plus petite que dans le cas du fleuve Delaware.

En procédant comme ci-dessus, on applique les tests d'ajustement du khi-deux pour chaque distribution considérée de la variable aléatoire T . Les résultats sont présentés au tableau (3.4). La figure (3.2) illustre les fonctions de répartition cumulatives. On remarque ici que les distributions exponentielle et Rayleigh sont rejetées, alors que la distribution lognormale donne clairement le meilleur ajustement aux données.

Les valeurs des quatre critères utilisés pour comparer les modèles sont présentées dans le tableau (3.5). On néglige la distribution de Rayleigh, parce qu'elle ne s'ajuste pas bien aux données. Les valeurs des critères Nash et Corr sont presque similaires comme dans le cas du fleuve Delaware, alors que les valeurs des critères de Pc et Mape sont beaucoup plus petites pour le PRF que pour le PPF.

Enfin, les valeurs observées et prévues des fleuves Delaware et Hudson aux instants $t + 1$ et $t + 2$ sont illustrées aux figures (3.3), (3.4), (3.5) et (3.6) respectivement, basées sur le PPF et le PRF avec la distribution lognormale pour le temps d'inter-arrivée T . On note que le

PRF est généralement meilleur que le PPF au niveau des valeurs prévues extrêmes du débit, ce qui répond à notre objectif principal.

Tableau 3.4 Tests d'ajustement de la distribution de la variable aléatoire T dans le cas du fleuve Hudson

Distributions	Exponentielle	Lognormale	Gamma	Weibull	Rayleigh
Paramètres	0,1198	1,8754 ; 0,6625	9,2134 ; 1,3399	2,1785 ; 3,8324	7,9978
Valeurs p	0,0013	0,6983	0,3203	0,2532	0,0001

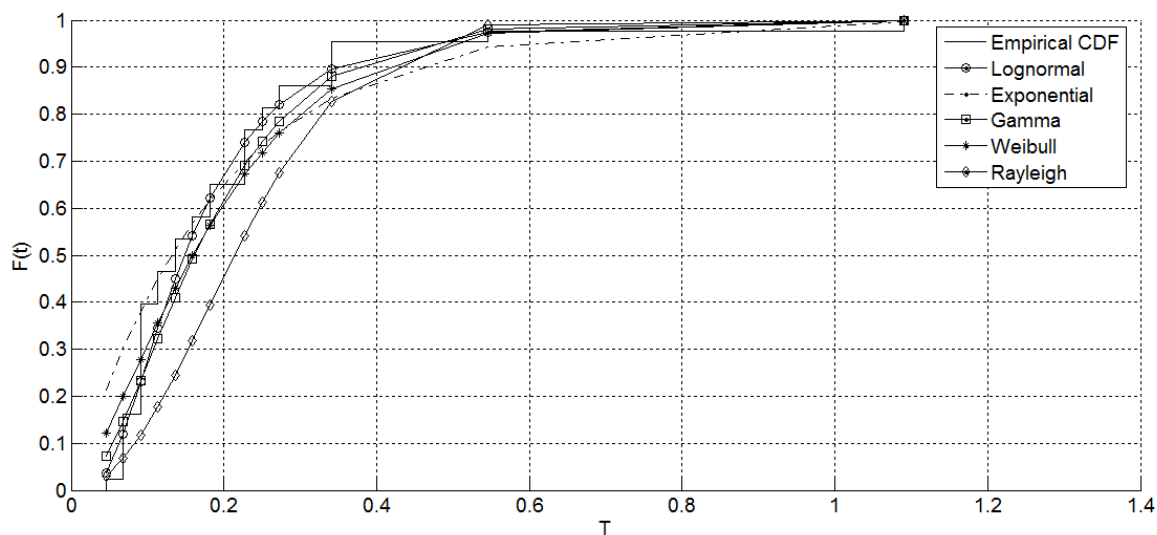


Figure 3.2 Les fonctions de répartition empirique et ajustée de la variable aléatoire T dans le cas du fleuve Hudson

Tableau 3.5 Critères de comparaison pour les modèles de prévisions dans le cas du fleuve Hudson

		Processus de renouvellement filtré			Processus de Poisson filtré
		Lognormale	Gamma	Weibull	Exponentielle
$X(t + 1)$	Mape	12,33%	12,14%	12,33%	15,58%
	Nash	0,8240	0,8243	0,8240	0,8360
	Pc	0,2875	0,2895	0,2872	0,4818
	Corr	0,9146	0,9146	0,9146	0,9146
$X(t + 2)$	Mape	20,84%	21,00%	21,15%	32,78%
	Nash	0,5841	0,5838	0,5835	0,6222
	Pc	0,4558	0,4533	0,4518	0,5575
	Corr	0,7908	0,7909	0,7910	0,7910

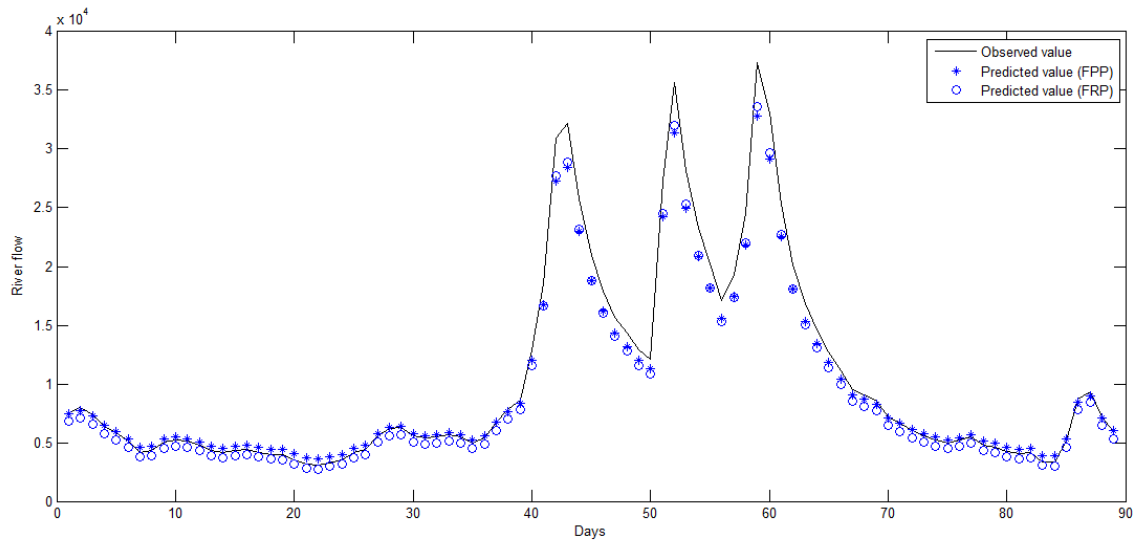


Figure 3.3 Les valeurs observées et prévues du fleuve Delaware à l'instant $t + 1$.

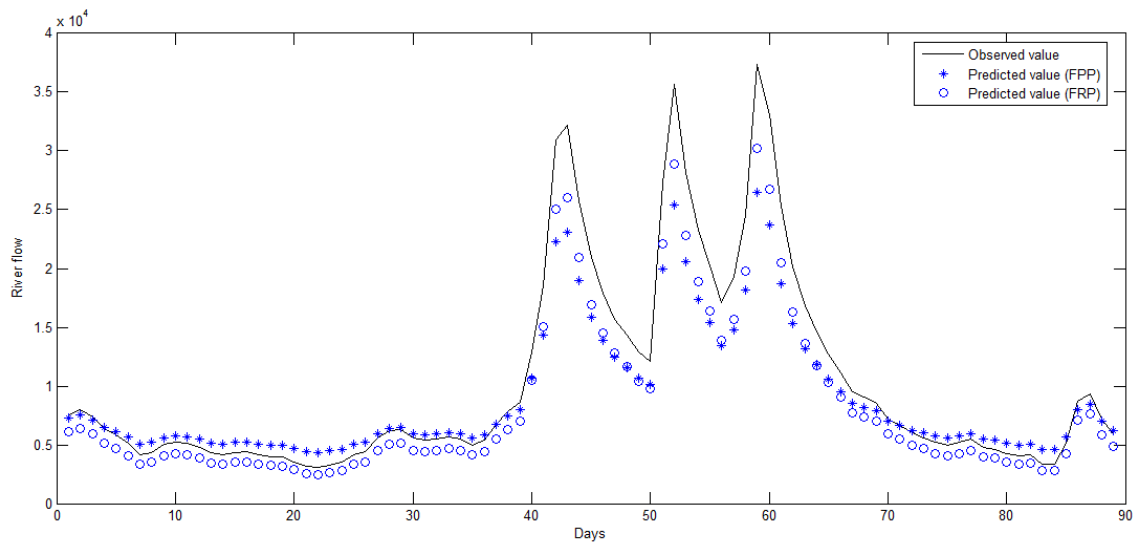


Figure 3.4 Les valeurs observées et prévues du fleuve Delaware à l'instant $t + 2$.

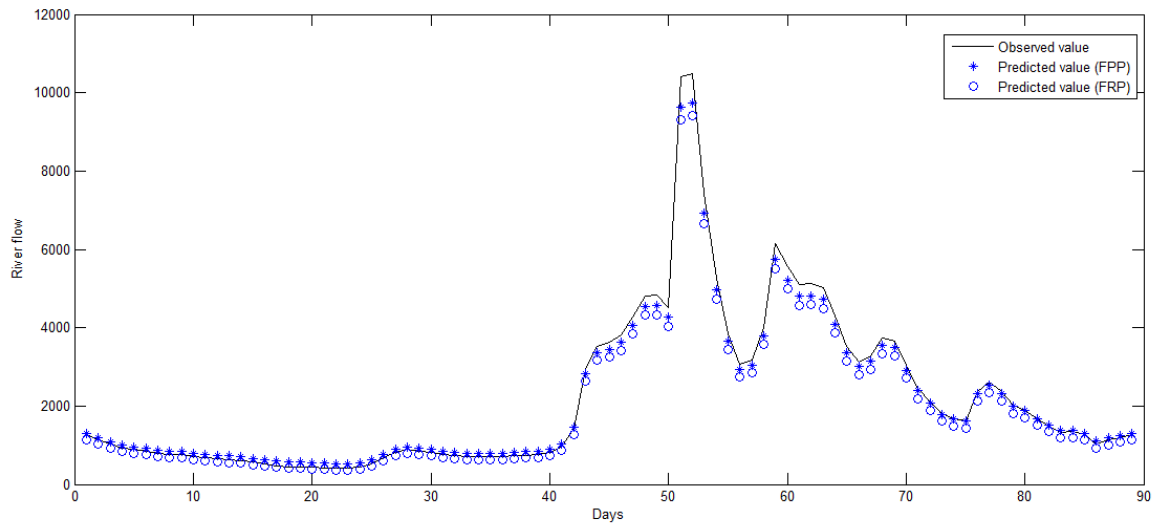


Figure 3.5 Les valeurs observées et prévues du fleuve Hudson à l'instant $t + 1$.

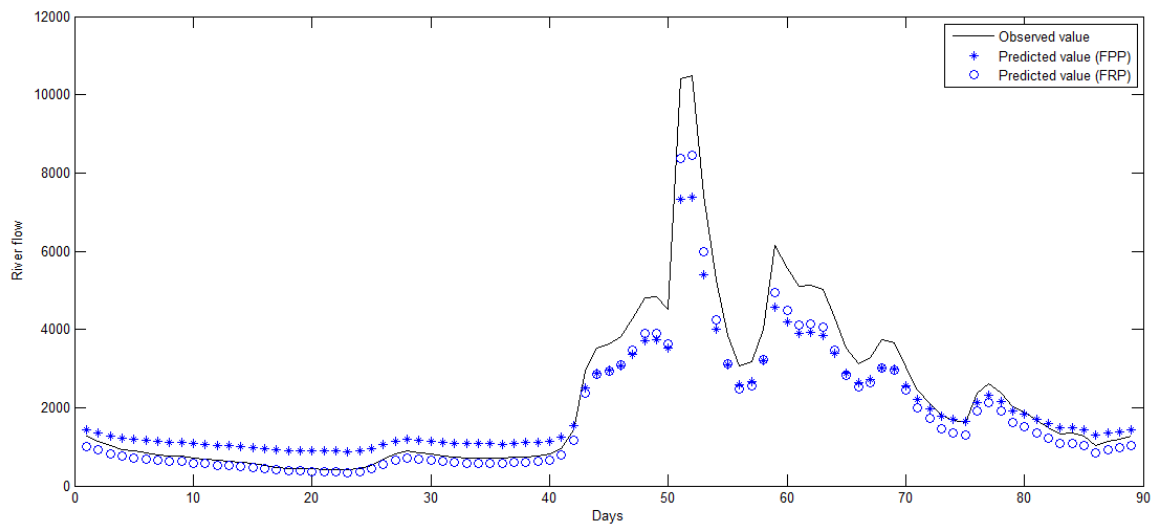


Figure 3.6 Les valeurs observées et prévues du fleuve Hudson à l'instant $t + 2$.

CHAPITRE 4 LA PROBABILITÉ THÉORIQUE DE DÉPASSEMENT D'UN SEUIL DE DÉBIT DONNÉ

On développe dans ce qui suit une formule pour la valeur du débit d'un cours d'eau à l'instant $t + 1$, étant donnée sa valeur observée à l'instant t . En premier lieu, dans le cas du processus de Poisson filtré $\{X(t), t \geq 0\}$ défini en (5.1), on peut écrire que

$$\begin{aligned}
 X(t+1) &= \sum_{n=1}^{N(t+1)} Y_n e^{-(t+1-\tau_n)/c} \\
 &= \sum_{n=1}^{N(t)} Y_n e^{-(t+1-\tau_n)/c} + \sum_{n=N(t^+)}^{N(t+1)} Y_n e^{-(t+1-\tau_n)/c} \\
 &= e^{-1/c} \sum_{n=1}^{N(t)} Y_n e^{-(t-\tau_n)/c} + \sum_{n=N(t^+)}^{N(t+1)} Y_n e^{-(t+1-\tau_n)/c} \\
 &= e^{-1/c} X(t) + \sum_{n=N(t^+)}^{N(t+1)} Y_n e^{-(t+1-\tau_n)/c}.
 \end{aligned} \tag{4.1}$$

De là, on obtient que

$$X(t+1) \mid \{X(t) = x\} = x e^{-1/c} + \sum_{n=N(t^+)}^{N(t+1)} Y_n e^{-(t+1-\tau_n)/c}. \tag{4.2}$$

Maintenant, on suppose qu'il y a au plus un événement du processus de Poisson qui arrive entre t et $t + 1$. On considère donc que la probabilité d'arrivée d'au moins deux événements dans l'intervalle $(t, t + 1]$ est négligeable. Il s'ensuit que

$$X(t+1) \mid \{X(t) = x\} \simeq \begin{cases} x e^{-1/c} & \text{avec une probabilité de } e^{-\lambda}, \\ x e^{-1/c} + Y e^{-(t+1-\tau)/c} & \text{avec une probabilité de } \lambda e^{-\lambda}, \end{cases}$$

où Y est une variable aléatoire distribuée comme Y_n et τ est uniformément distribuée sur $(t, t + 1]$. Effectivement, comme il est bien connu, sachant qu'un événement d'un processus de Poisson s'est produit dans un intervalle donné $[a, b]$, l'instant d'arrivée de cet événement est uniformément distribué sur $[a, b]$ (Lefebvre (2007)).

Remarque : En pratique, les valeurs de débit sont généralement enregistrées sur une base quotidienne. Par conséquent, il n'est pas vraiment possible de déterminer le nombre d'événements

ments qui se sont produits dans l'intervalle $(t, t+1]$. On suppose alors qu'un événement a eu lieu lorsque la valeur de $X(t+1)$ est supérieure ou égale à $X(t)$. En outre, on déterminera plus loin dans l'application, avec l'estimation ponctuelle de λ qui sera calculée, que la probabilité de deux événements ou plus au cours d'une journée est environ égale à 0,03, de sorte que l'on peut négliger cette possibilité.

Soit

$$p(x, y) := P[X(t+1) \geq x+y \mid \{X(t) = x\}]. \quad (4.3)$$

On veut estimer cette probabilité lorsque x est grand, de sorte que si y aussi est grand, il y aura certainement un risque de crue. Puisque $X(t+1)$ sera inférieur à $X(t)$ s'il n'y a pas d'événement entre t et $t+1$ (selon le modèle), on déduit de ce qui précède qu'on peut écrire que

$$\begin{aligned} p(x, y) &\simeq P[x e^{-1/c} + Y e^{-(t+1-\tau)/c} \geq x+y] \lambda e^{-\lambda} \\ &= \lambda e^{-\lambda} \int_t^{t+1} P[Y e^{-(t+1-s)/c} \geq x(1 - e^{-1/c}) + y] ds \\ &= \lambda e^{-\lambda} \int_0^1 P[Y e^{-(1-u)/c} \geq x(1 - e^{-1/c}) + y] du \\ &= \lambda e^{-\lambda} \int_0^1 P[Y \geq [x(1 - e^{-1/c}) + y] e^{(1-u)/c}] du \\ &= \lambda e^{-\lambda} \int_0^1 \exp \left\{ -\mu [x(1 - e^{-1/c}) + y] e^{(1-u)/c} \right\} du, \end{aligned} \quad (4.4)$$

où Y suit une distribution exponentielle de paramètre μ .

L'intégrale indéfinie de la probabilité $p(x, y)$ peut être exprimée comme suit :

$$\int \exp \left\{ -\mu [x(1 - e^{-1/c}) + y] e^{(1-u)/c} \right\} du = -c \operatorname{Ei} \left(-e^{\frac{1-u}{c}} \mu (y - e^{-1/c} x + x) \right), \quad (4.5)$$

où $\operatorname{Ei}(x)$ est l'exponentielle intégrale.

En second lieu, on dérive une formule qui correspond à celle en (4.4) lorsqu'on utilise le processus de diffusion avec sauts $\{X_D(t), t \geq 0\}$, défini en (5.10), pour modéliser les variations des débits :

$$X_D(t) = D(t) + \sum_{n=1}^{N(t)} Y_n \quad \text{pour } t \geq 0. \quad (4.6)$$

On choisit pour le processus de diffusion $\{D(t), t \geq 0\}$ un processus de Wiener (ou mouvement brownien) avec un paramètre de dérive θ (négatif) et un paramètre de dispersion σ^2 :

$$D(t) = \theta t + \sigma B(t), \quad (4.7)$$

où $\{B(t), t \geq 0\}$ est un mouvement brownien standard. $B(t) \sim N(0, t)$ et $B(0) = 0$, ce qui implique que $D(0) = 0$. D'où, on peut écrire que

$$D(t) \sim N(\theta t, \sigma^2 t). \quad (4.8)$$

C'est-à-dire que $D(t)$ a une distribution gaussienne de moyenne θt et de variance $\sigma^2 t$.

Si au plus un signal se produit dans l'intervalle $(t, t + 1]$, le processus de diffusion avec sauts à l'instant $t + 1$ est donné par

$$X_D(t + 1) = D(t + 1) + \sum_{n=1}^{N(t+1)} Y_n \quad \text{pour } t \geq 0. \quad (4.9)$$

Contrairement au cas du processus de Poisson filtré, il est possible que $X_D(t+1)$ soit supérieur à $X_D(t)$ même s'il n'y a pas d'événement du processus de Poisson dans l'intervalle $(t, t + 1]$.

S'il n'y a aucun signal qui se produit, avec une probabilité de $e^{-\lambda}$, alors

$$X_D(t + 1) | \{X_D(t) = x\} \sim N(x + \theta, \sigma^2). \quad (4.10)$$

S'il y a un signal qui se produit, avec une probabilité de $\lambda e^{-\lambda}$, alors

$$X_D(t + 1) | \{X_D(t) = x\} = Z_1 + Y, \quad (4.11)$$

où $Z_1 \sim N(x + \theta, \sigma^2)$ et $Y \sim \text{Exp}(\mu)$.

On obtient alors la formule suivante :

$$\begin{aligned} p_D(x, y) &:= P[N(\theta, \sigma^2) \geq y] e^{-\lambda} + P[N(\theta, \sigma^2) + Y \geq y] \lambda e^{-\lambda} \\ &= e^{-\lambda} Q\left(\frac{y - \theta}{\sigma}\right) + \lambda e^{-\lambda} \int_{-\infty}^{\infty} P[Y \geq y - z | Z = z] f_Z(z) dz, \end{aligned} \quad (4.12)$$

où $Q(z) := P[Z > z]$ et $Z \sim N(\theta, \sigma^2)$.

De façon plus explicite,

$$p_D(x, y) = e^{-\lambda} Q\left(\frac{y - \theta}{\sigma}\right) + \lambda e^{-\lambda} e^{-\mu y} \int_{-\infty}^{\infty} e^{\mu z} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(z - \theta)^2}{2\sigma^2}\right\} dz. \quad (4.13)$$

On note que cette probabilité ne dépend pas de x , contrairement à $p(x, y)$. Cependant, les paramètres du modèle dépendent de x .

L'intégrale indéfinie de la probabilité $p_D(x, y)$ peut être exprimée au moyen de la fonction d'erreur **erf** telle que :

$$\int e^{\mu z} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(z-\theta)^2}{2\sigma^2}\right\} dz = -\frac{1}{2} e^{\frac{\mu^2\sigma^2}{2} + \mu\theta} \operatorname{erf}\left(\frac{\mu\sigma^2 + \theta - z}{\sqrt{2}\sigma}\right). \quad (4.14)$$

Finalement, on utilise aussi la régression linéaire pour estimer la probabilité qui correspond à $p(x, y)$. Soit

$$X_R(t+1) = aX_R(t) + b + \epsilon, \quad (4.15)$$

où $\epsilon \sim N(0, \sigma^2)$. Alors, la formule en (4.4) devient simplement

$$p_R(x, y) = P[N(ax + b, \sigma^2) \geq x + y] = Q\left(\frac{(1-a)x + y - b}{\sigma}\right). \quad (4.16)$$

4.1 Application et résultats

4.1.1 Implémentation des modèles de distribution de probabilités de dépassement

On ne considère ici que le fleuve Delaware pour l'application des résultats du chapitre (4) (bien livrable 2). On applique alors les trois modèles considérés, le processus de Poisson filtré (PPF), le processus de diffusion avec sauts (PDS) et la régression linéaire pour estimer la probabilité que le débit augmente d'au moins y pieds cubes par seconde en un jour, pour différentes valeurs de y , lorsque sa valeur courante est égale à x .

On utilise les données à partir du 23 janvier 2003 au 22 janvier 2014, c'est-à-dire une période de 11 ans. Après avoir estimé les paramètres, on peut calculer les estimations ponctuelles des probabilités $p(x, y)$, $p_D(x, y)$ et $p_R(x, y)$. Finalement, ces estimations seront comparées aux fréquences observées des événements considérés.

Rappelons que les paramètres qu'on doit estimer dépendent de la valeur du débit x . On va calculer les estimations en question pour $x = 10000 \text{ pi}^3/\text{s}$. On choisit cette valeur parce qu'elle représente une valeur du débit qui est déjà élevée, et qui est telle que si le débit augmente, par exemple, d'au moins $5000 \text{ pi}^3/\text{s}$ dans une journée, alors il peut y avoir des risques d'inondation. Par ailleurs, on a besoin de suffisamment d'observations pour obtenir des fréquences fiables. En fait, il n'y a pas suffisamment de journées pendant lesquelles le débit courant est exactement égal à $10000 \text{ pi}^3/\text{s}$ dans la série de données. Par conséquent, on a décidé d'utiliser tous les cas où x appartient à l'intervalle $[10000, 11000)$. Il y a 106 données.

Avant l'estimation des paramètres, on doit vérifier si la distribution exponentielle est effective-

ment un modèle approprié pour les variables aléatoires T_n , ce qui est le cas si $\{N(t), t \geq 0\}$ est un processus de Poisson. En supposant que nos données constituent un échantillon aléatoire particulier de la distribution exponentielle, on vérifie cette hypothèse par le test d'ajustement du khi-deux. Étant donné que la moyenne des observations est égale à 4,708, le paramètre de la distribution exponentielle est égale à $1/4,708$. Les résultats sont présentés dans le tableau (4.1). On trouve que la statistique du test est égale à 6,07, ce qui donne, comparativement à la valeur de la distribution de khi-deux à 6 degrés de liberté, une valeur p de 0,415. D'où on peut conclure que la distribution exponentielle est certainement un modèle acceptable pour les variables aléatoires T_n .

Tableau 4.1 Test d'ajustement pour la distribution des variables aléatoires T_n

Intervalles	(0; 2]	(2; 4]	(4; 6]	(6; 8]	(8; 10]	(10; 12]	(12; 14]	(14; ∞]
Observées	39	20	21	13	5	4	1	3
Probabilités	0,346	0,226	0,148	0,097	0,063	0,041	0,027	0,051
Espérées	36,7	24,0	15,7	10,3	6,7	4,3	2,9	5,4

4.1.2 Estimation des paramètres

D'abord, on estime les paramètres du processus de Poisson filtré $\{X(t), t \geq 0\}$, qui sont la constante c de la fonction de réponse, le taux λ du processus de Poisson $\{N(t), t \geq 0\}$, et le paramètre μ des variables aléatoires Y_n de distribution exponentielle.

Selon le modèle, lorsqu'il n'y a pas d'événement dans l'intervalle $(t, t+1]$, on peut estimer la constante c selon l'équation (4.17), en considérant tous les cas où la valeur de $X(t+1)$ est inférieure à celle de $X(t)$ et en calculant la moyenne arithmétique du quotient $R = \frac{X(t+1)}{X(t)}$:

$$\hat{c} = -\frac{1}{\ln \bar{R}}. \quad (4.17)$$

Il y a 78 observations (parmi les 106) dans la série de données pour lesquelles $X(t+1) < X(t)$. Avec ces observations, on trouve $\hat{c} \simeq 7,72$.

Ensuite, puisque la valeur du débit a augmenté (ou est demeurée stable) 28 fois parmi 106, dans la période considérée, on peut estimer le taux λ du processus de Poisson $\{N(t), t \geq 0\}$ par

$$\hat{\lambda} = \frac{28}{106}. \quad (4.18)$$

Finalement, pour estimer le paramètre μ de la distribution exponentielle, on utilise le fait que lorsqu'un événement arrive entre t et $t + 1$, il se produit en moyenne à l'instant $t + 1/2$. On déduit alors du modèle que la valeur approximative du signal est donnée par

$$Y \simeq \frac{X(t+1) - e^{-1/c} X(t)}{e^{-1/2c}}. \quad (4.19)$$

En calculant (avec $c \simeq 7,72$) la moyenne arithmétique des 28 observations pour lesquelles $X(t+1) \geq X(t)$, on obtient que la valeur moyenne des signaux est environ 8465. Il s'ensuit que $\mu = 1/8465$. Alors, l'estimation de la probabilité $p(10000, y)$ est

$$\hat{p}(10000, y) \simeq 0,264 e^{-0,264} \int_0^1 \exp \left\{ - \left(\frac{1}{8465} \right) [10000(1 - e^{-1/7,72}) + y] e^{(1-u)/7,72} \right\} du. \quad (4.20)$$

Dans le cas du processus de diffusion avec sauts $\{X_D(t), t \geq 0\}$, on doit estimer le paramètre de dérive θ et le paramètre de dispersion σ^2 du processus de Wiener $\{D(t), t \geq 0\}$, le paramètre μ de la distribution exponentielle et le taux λ du processus de Poisson. On peut utiliser la même estimation de λ que celle calculée pour le processus de Poisson filtré : $\hat{\lambda} = 28/106 \simeq 0,2644$.

Lorsqu'il n'y a pas d'événement dans l'intervalle $(t, t + 1]$, la valeur du débit se déplace, en moyenne, de $X_D(t)$ à $X_D(t + 1) = X_D(t) + \theta$. Donc, en calculant la moyenne arithmétique de la décroissance des 78 fois où $X_D(t + 1) < X_D(t)$, on trouve que

$$\hat{\theta} \simeq -1269,1. \quad (4.21)$$

De façon similaire, on peut estimer σ^2 en calculant la variance des 78 décroissances du débit. On obtient que

$$\hat{\sigma}^2 \simeq (662,6)^2. \quad (4.22)$$

Pour estimer le paramètre μ de la distribution exponentielle des variables aléatoires Y_n , on considère les 28 cas où le débit à l'instant $t + 1$ est supérieur ou égal au débit à l'instant t . Étant donné que lorsqu'un événement arrive dans l'intervalle $(t, t + 1]$, on peut écrire que (en moyenne)

$$X_D(t + 1) = X_D(t) + \theta + Y \quad (4.23)$$

où $Y \sim \text{Exp}(\mu)$, pour estimer μ on peut calculer l'augmentation moyenne du débit et soustraire $\hat{\theta}$ de cette moyenne. On trouve que l'augmentation moyenne est égale à 6664, ainsi on pose que

$$\hat{\mu} \simeq \frac{1}{7933}. \quad (4.24)$$

Alors, on obtient que l'estimation de la probabilité $p_D(10000, y)$ est donnée par

$$\begin{aligned} \hat{p}_D(10000, y) &= e^{-0,264} Q\left(\frac{y + 1269, 1}{662, 6}\right) \\ &+ 0,264 e^{-0,264} e^{-(1/7933)y} \\ &\times \int_{-\infty}^{\infty} e^{-(1/7933)z} \frac{1}{\sqrt{2\pi}(662, 6)} \exp\left\{-\frac{(z + 1269, 1)^2}{2(662, 6)^2}\right\} dz. \end{aligned} \quad (4.25)$$

Enfin, on considère le modèle de régression (4.15), on estime les paramètres a , b et σ^2 et on trouve que

$$X_R(t + 1) = 2,61X_R(t) - 16031$$

et $\hat{\sigma} \simeq 6180,49$. Il s'ensuit que

$$\hat{p}_R(10000, y) = Q\left(\frac{(-1, 61)10000 + y + 16031}{6180,49}\right) = Q\left(\frac{y - 69}{6180,49}\right). \quad (4.26)$$

4.1.3 Résultats numériques

On applique les formules précédentes avec $y = 5000$, 10000 et 20000 . Les résultats sont présentés dans le tableau (4.2), avec les fréquences observées correspondantes. D'après ces résultats, on note que $\hat{p}(10000, y)$ et $\hat{p}_D(10000, y)$ sont pratiquement égales, même si les modèles correspondants sont différents. En effet, dans le cas du PPF (5.1), on utilise une fonction de réponse telle que le débit décroît exponentiellement (de façon déterministe) entre les événements du processus de Poisson, alors que dans le cas du PDS (5.10) le débit évolue selon un processus de Wiener entre les événements.

Par ailleurs, on peut affirmer que $\hat{p}(10000, y)$ et $\hat{p}_D(10000, y)$ sont de bonnes approximations des fréquences correspondantes. Ceci est encore plus impressionnant lorsqu'on se rappelle que les fréquences ont été calculées en utilisant toutes les valeurs du débit comprises dans l'intervalle $[10000, 11000)$ plutôt que de considérer seulement les observations égales exactement à 10000 , parce que la taille de l'échantillon de données aurait alors été trop petite.

Finalement, il est clair que le modèle de régression linéaire ne fournit pas des estimations précises des probabilités considérées, au moins lorsqu'on compare les estimations obtenues aux fréquences correspondantes. Ceci est plutôt surprenant, parce que Lefebvre (2003) a trouvé qu'on peut obtenir de bonnes prévisions à court terme des débits de cours d'eau en utilisant la régression linéaire simple.

Tableau 4.2 Valeurs numériques des différentes estimations, et fréquences observées correspondantes

y	$\hat{p}(10000, y)$	$\hat{p}_D(10000, y)$	$\hat{p}_R(10000, y)$	Fréquences
5000	0.0926	0.0923	0.2125	0.0943
10000	0.0493	0.0492	0.0540	0.0472
20000	0.0140	0.0140	0.0006	0.0189

CHAPITRE 5 LA DISTRIBUTION DU DÉBIT AU MOMENT DE LA PROCHAINE AUGMENTATION

À partir du processus de Poisson filtré défini en (5.1) :

$$X(t) = \sum_{n=1}^{N(t)} Y_n e^{-(t-\tau_n)/c} \quad (X(t) = 0 \text{ si } N(t) = 0), \quad (5.1)$$

on déduit que

$$X(\tau_{n+1}) = \sum_{k=1}^{N(\tau_{n+1})} Y_k e^{-(\tau_{n+1}-\tau_k)/c} = \sum_{k=1}^{n+1} Y_k e^{-(\tau_{n+1}-\tau_k)/c}. \quad (5.2)$$

Puisque

$$X(\tau_n) = \sum_{k=1}^{N(\tau_n)} Y_k e^{-(\tau_n-\tau_k)/c} = \sum_{k=1}^n Y_k e^{-(\tau_n-\tau_k)/c}, \quad (5.3)$$

on peut écrire que

$$\begin{aligned} X(\tau_{n+1}) | X(\tau_n) &\simeq \sum_{k=1}^{n+1} Y_k e^{-(\tau_{n+1}-\tau_k)/c} | \sum_{k=1}^n Y_k e^{-(\tau_n-\tau_k)/c} \\ &\simeq \sum_{k=1}^n Y_k e^{-(\tau_{n+1}-\tau_k)/c} + Y_{n+1} e^{-(\tau_{n+1}-\tau_{n+1})/c} | \sum_{k=1}^n Y_k e^{-(\tau_n-\tau_k)/c} \\ &\simeq \sum_{k=1}^n Y_k e^{[-(\tau_n-\tau_k)+\tau_n-\tau_{n+1}]/c} + Y_{n+1} \\ &\simeq \sum_{k=1}^n Y_k e^{-(\tau_n-\tau_k)/c} e^{-(\tau_{n+1}-\tau_n)/c} + Y_{n+1}. \end{aligned} \quad (5.4)$$

D'où

$$X(\tau_{n+1}) | X(\tau_n) \simeq X(\tau_n) e^{-(\tau_{n+1}-\tau_n)/c} + Y_{n+1}. \quad (5.5)$$

Selon le modèle défini en (5.1), le temps d'inter-arrivée $T := \tau_{n+1} - \tau_n$ est supposé suivre une distribution exponentielle de paramètre λ . Pour obtenir alors la distribution de la variable aléatoire conditionnelle $X(\tau_{n+1}) | X(\tau_n)$, on doit trouver un modèle pour l'amplitude des signaux Y . Comme on l'a mentionné précédemment, en hydrologie il est souvent supposé que Y est une variable aléatoire exponentielle de paramètre μ . Si ces distributions sont acceptables, on doit calculer la convolution des variables aléatoires $X(\tau_n) e^{-T/c}$ et Y pour une valeur fixe x de $X(\tau_n)$. On doit aussi estimer le paramètre c (≥ 0) du modèle.

Si $T \sim \text{Exp}(\lambda)$, on trouve que la distribution de probabilité de $R := x e^{-T/c}$ est donnée par

$$\begin{aligned}
 P(R \leq r) &= P\left(x e^{-T/c} \leq r\right) \\
 &= P\left(T > -c \ln\left(\frac{r}{x}\right)\right) \\
 &= e^{-\lambda c \left\{\ln\left(\frac{x}{r}\right)\right\}} \\
 &= \left(\frac{r}{x}\right)^{\lambda c} \text{ pour } r \leq x.
 \end{aligned} \tag{5.6}$$

La fonction densité de probabilité de R est alors donnée par

$$f_R(r) = \frac{\lambda c}{r} \left(\frac{r}{x}\right)^{\lambda c} \text{ pour } r \leq x. \tag{5.7}$$

D'où, si $Y \sim \text{Exp}(\mu)$, la fonction densité de $V(x) := X(\tau_{n+1}) \mid \{X(\tau_n) = x\}$ peut être écrite comme suit :

$$\begin{aligned}
 f_{V(x)}(v) &:= \int_{-\infty}^{\infty} f_R(r) f_Y(v-r) dr \\
 &= \int_{-\infty}^v f_R(r) f_Y(v-r) dr \\
 &= \int_0^{\min(x,v)} f_R(r) f_Y(v-r) dr \\
 &= \int_0^{\min(x,v)} \frac{\lambda c}{r} \left(\frac{r}{x}\right)^{\lambda c} \mu e^{-\mu(v-r)} dr.
 \end{aligned} \tag{5.8}$$

Cette formule est valable pour $v \geq 0$.

L'intégrale indéfinie de $f_{V(x)}(v)$ est donnée par :

$$\begin{aligned}
 \int f_R(r) f_Y(v-r) dr &= -c \lambda \mu e^{-\mu v} (-\mu r)^{-c\lambda} \left(\frac{r}{x}\right)^{c\lambda} \Gamma(c\lambda, -\mu r) \\
 &= -c \lambda \mu e^{-\mu v} \left(\frac{r}{x}\right)^{c\lambda} E_{1-c\lambda}(-\mu r)
 \end{aligned} \tag{5.9}$$

où $\Gamma(a, x)$ est la fonction Gamma incomplète et $E_n(x)$ est l'intégrale exponentielle.

En utilisant le processus de diffusion avec sauts $\{X_D(t), t \geq 0\}$ défini par (5.10), avec le processus de diffusion $\{D(t), t \geq 0\}$ défini par le processus de Wiener (avec $D(0)=0$) de paramètres de dérive θ et de dispersion σ^2 :

$$X_D(t) = D(t) + \sum_{n=1}^{N(t)} Y_n \text{ pour } t \geq 0, \tag{5.10}$$

la fonction de densité de la variable aléatoire $V(x)$ est donnée par :

$$\begin{aligned} f_{V(x)}(v) &= \lim_{dv \downarrow 0} \frac{1}{dv} P \left[\left\{ D(\tau_{n+1}) + \sum_{k=1}^{n+1} Y_k \right\} \in (v, v + dv) \mid \left\{ D(\tau_n) + \sum_{k=1}^n Y_k = x \right\} \right] \\ &= \lim_{dv \downarrow 0} \frac{1}{dv} P [\{D(\tau_{n+1}) - D(\tau_n) + Y_{n+1}\} \in (v - x, v - x + dv)]. \end{aligned} \quad (5.11)$$

Puisque $\{D(t), t \geq 0\}$ a des accroissements indépendants et stationnaires, on peut affirmer que $D(\tau_{n+1}) - D(\tau_n)$ a la même distribution que $D(T)$ (avec $D(0) = 0$), soit la distribution gaussienne de moyenne θT et de variance $\sigma^2 T$. Il s'ensuit, en conditionnant sur T et en dénotant $D(T) \mid \{T = t\}$ par S , que

$$\begin{aligned} f_{V(x)}(v) &= \int_0^\infty \lim_{dv \downarrow 0} \frac{1}{dv} P[\{S + Y\} \in (v - x, v - x + dv)] f_T(t) dt \\ &= \int_0^\infty \int_{-\infty}^\infty f_S(s) f_Y(v - x - s) f_T(t) ds dt \\ &= \int_0^\infty \int_{-\infty}^{v-x} f_S(s) f_Y(v - x - s) f_T(t) ds dt \\ &= \int_0^\infty \int_{-\infty}^{v-x} \frac{1}{\sqrt{2\pi\sigma t^{1/2}}} \exp \left\{ -\frac{(s - \theta t)^2}{2\sigma^2 t} \right\} \mu e^{-\mu(v-x-s)} \lambda e^{-\lambda t} ds dt. \end{aligned} \quad (5.12)$$

5.1 Application et résultats

5.1.1 Distribution de $X(\tau_{n+1}) \mid X(\tau_n)$

Rappelons que dans le chapitre 4 (bien livrable 2), les paramètres du modèle dépendent des valeurs des débits. En effet, la valeur du taux λ du processus de Poisson, en particulier, est probablement plus élevée lorsque le débit $X(t)$ est très grand que lorsqu'il est faible.

On garde les mêmes estimations des paramètres c , θ et σ : $\hat{c} = 7,72$, $\hat{\theta} = -1269,1$ et $\hat{\sigma} = 662,6$. Cependant, on ne peut pas utiliser les mêmes estimations des paramètres λ et μ obtenues précédemment (au chapitre 4), car dans ce cas, non seulement $X(t)$ doit être dans l'intervalle $[10000, 11000)$, la valeur de $X(t - 1)$ doit être inférieure ou égale à $X(t)$. En effet, puisqu'on s'intéresse à trouver la distribution de $X(\tau_{n+1})$, sachant $X(\tau_n)$, l'observation courante correspond à une valeur de t à laquelle un événement se produit. En fait, dans le cas du processus de diffusion avec sauts, la valeur de $X(\tau_n)$ peut être inférieure à $X(\tau_n - 1)$ même si un événement se produit à l'instant $t = \tau_n$ (à cause de la part de diffusion du processus). Cependant, puisque θ est en pratique une grande constante négative, on peut supposer que $X(\tau_n) \geq X(\tau_n - 1)$.

Avec cette contrainte additionnelle, la taille de l'échantillon de données est réduite à 30. En

procédant comme le chapitre précédent, on obtient les estimations suivantes des paramètres λ et μ : $\hat{\lambda} = 1/3,8$ (3,8 est la moyenne arithmétique des observations de la variable aléatoire T) et $\hat{\mu} = 1/5010$.

Dans le chapitre précédent, on a rappelé le fait que si un événement arrive dans l'intervalle $(t, t + 1]$, l'instant d'arrivée de cet événement est uniformément distribué dans cet intervalle. Alors, on peut écrire que

$$\frac{1}{\hat{\mu}} = \frac{1}{30} \sum_{n=1}^{30} \frac{X(\tau_{n+1}) - e^{-1/c} X(\tau_n)}{e^{-1/(2c)}} \quad (5.13)$$

(avec la constante c remplacée par sa valeur estimée $\hat{c} = 7,72$).

Remarque : puisque la variable aléatoire T est discrétisée, on peut négliger le terme $e^{-1/(2\hat{c})} \simeq 1,067$ et estimer le paramètre μ en considérant les 30 journées pour lesquelles il y a une augmentation de la valeur du débit, et on calcule la moyenne arithmétique des différences $X(\tau_{n+1}) - e^{-1/c} X(\tau_n)$ pour $n = 1, \dots, 30$. Puis, on pose μ égal à l'inverse de la moyenne arithmétique.

De même, dans le cas du processus de diffusion avec sauts, s'il y a un événement dans l'intervalle $(t, t + 1]$, alors $X_D(t + 1) = X_D(t) + \theta + Y$ (4.23). Il s'ensuit qu'on peut estimer μ en posant

$$\frac{1}{\hat{\mu}} = \frac{1}{30} \sum_{n=1}^{30} \{X(\tau_{n+1}) - [X(\tau_n) + \theta]\}. \quad (5.14)$$

On obtient une estimation de μ légèrement différente. Mais, pour des raisons de simplicité, on a décidé d'utiliser la valeur de $\hat{\mu}$ calculée ci-dessus (5.13). Encore une fois, cela ne fera pas une différence significative dans l'ajustement du modèle aux données.

On peut maintenant tester l'hypothèse que les variables aléatoires T et Y suivent une distribution exponentielle. On effectue des tests d'ajustement du khi-deux. Tout d'abord, dans le cas de $T \sim \text{Exp}(\hat{\lambda} = 1/3,8)$, on obtient les résultats présentés dans le tableau (5.1). La statistique du test est égale à 6,7, ce qui correspond à une valeur p d'environ 0,15.

Tableau 5.1 Test d'ajustement pour la distribution de la variable aléatoire T

Intervalles	(0; 2]	(2; 4]	(4; 6]	(6; 8]	(8; 10]	(10; ∞)
Observées	17	4	1	4	2	2
Probabilités	0,41	0,24	0,14	0,084	0,050	0,072
Espérées	12,3	7,2	4,2	2,52	1,50	2,16

Par conséquent, le modèle exponentiel ne peut pas être rejeté. En outre, parce que le test est sensible à la sélection des sous-intervalles, on a effectué un second essai avec un choix différent de ces sous-intervalles. On voit dans le tableau (5.2) en annexe A, que l'ajustement de la distribution aux données est meilleur que le précédent. La statistique du test est réduite à 2,54, avec une valeur p correspondante de 0,28.

Remarque : Parce que la variable aléatoire T est discrétisée, on pourrait de façon équivalente tester l'hypothèse que T a une distribution géométrique (commençant à 1) de paramètre $1 - e^{-\hat{\lambda}}$.

Tableau 5.2 Second test d'ajustement pour la distribution de la variable aléatoire T

Intervalles	(0, 3]	(3, 6]	(6, 9]	(9, ∞)
Observations	18	4	4	4
Probabilités	0.55	0.25	0.11	0.09
Espérances	16.5	7.5	3.3	2.7

Ensuite, on va tester l'hypothèse que $Y \sim \text{Exp}(1/5010)$. Comme ci-dessus, on a effectué deux tests avec différents sous-intervalles. Les résultats sont présentés dans les tableaux (5.3) et (5.4) en annexe A,. Les valeurs respectives de la statistique de test sont 4,18 et 2,44, ce qui correspond à des valeurs p de 0,243 et 0,295. Ainsi, on peut facilement accepter le modèle exponentiel pour la variable aléatoire Y .

De ce qui précède, on peut affirmer que les formules (5.8) et (5.12), avec les estimations ponctuelles calculées ci-dessus, pour la fonction de densité de $V(x)$ sont valables. La figure (5.1) présente les deux courbes correspondantes. La fonction de densité obtenue à partir de (5.12) est telle qu'il existe une probabilité non négligeable que $V(10000)$ prenne une valeur négative, ce qui bien sûr ne peut pas se produire en réalité. Par conséquent, le processus de Poisson filtré semble être un modèle légèrement meilleur pour cette variable aléatoire.

Tableau 5.3 Test d'ajustement pour la distribution de la variable aléatoire Y (en milliers de π^3/s)

Intervalles	(0; 3]	(3; 6]	(6; 9]	(9; 12]	(12; ∞)
Observées	18	3	4	2	3
Probabilités	0,451	0,248	0,136	0,0747	0,0912
Espérées	13,53	7,44	4,08	2,24	2,74

Tableau 5.4 Second test d'ajustement pour la distribution de la variable aléatoire Y (en milliers de π^3/s)

Intervalles	(0, 4]	(4, 8]	(8, 12]	(12, ∞)
Observations	20	4	3	3
Probabilités	0.55	0.25	0.11	0.09
Espérances	16.5	7.5	3.3	2.7

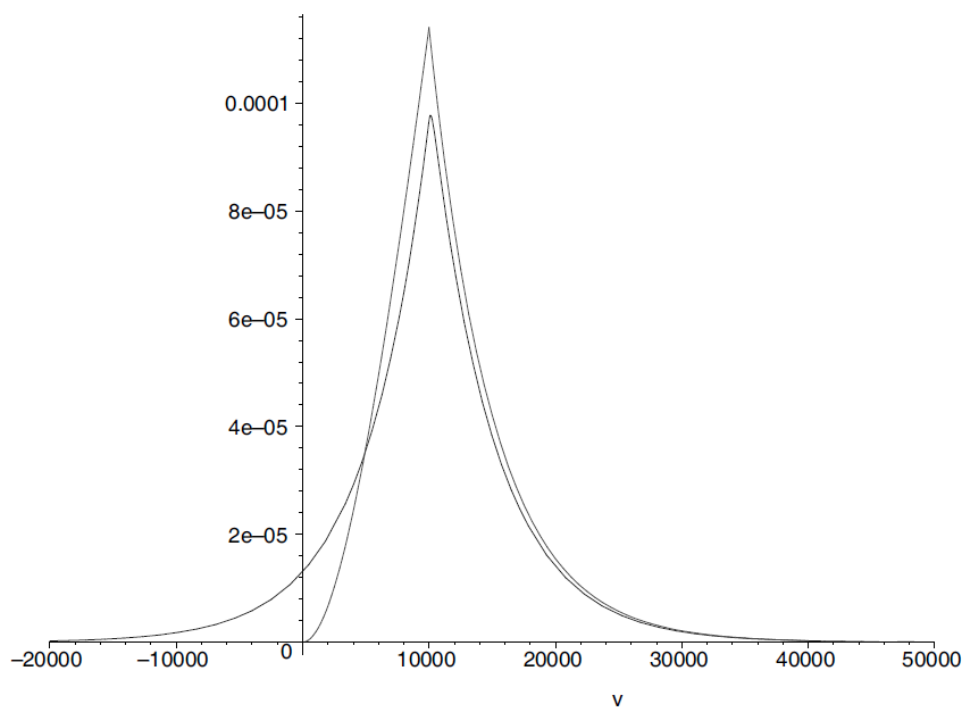


Figure 5.1 Les fonctions de densité de la v.a $V(10^4) = X(\tau_{n+1}) | \{X(\tau_n) = 10^4\}$ déduites de (5.8) (la courbe commençant de 0) et (5.12) avec les paramètres remplacés par leurs estimations ponctuelles.

On va également tester l'hypothèse que les fonctions de densité de probabilité issues du processus de Poisson filtré et du processus de diffusion avec sauts fournissent toutes les deux un bon ajustement aux données. Malheureusement, à cause du fait que la taille de l'échantillon de données est relativement faible (30), et qu'on a utilisé ces données pour estimer respectivement trois et quatre paramètres qui apparaissent dans les fonctions de densité, il est impossible d'obtenir un test très fiable. Néanmoins, on peut au moins essayer de vérifier si les modèles s'ajustent bien aux données. En regardant les tableaux (5.5) et (5.6) en annexe A,, on peut conclure que c'est bien le cas. Les statistiques des tests sont respectivement égales à 0,563 et 1,332, ce qui est significatif. Si on néglige le fait qu'on a estimé les paramètres du modèle, on constate que les valeurs p correspondantes sont de 0,755 et 0,514. Ainsi, les deux modèles sont appropriés pour la variable aléatoire $V(10000) = X(\tau_{n+1}) | \{X(\tau_n) = 10000\}$.

Tableau 5.5 Test d'ajustement du khi-deux de la fonction de densité (5.8) de la variable aléatoire $V(10000)$ (en milliers de π^3/s)

Intervalles	$(0, 10]$	$(10, 15]$	$(15, \infty)$
Observations	12	10	8
Probabilités	0.429	0.361	0.211
Espérances	12.87	10.83	6.33

Tableau 5.6 Test d'ajustement du khi-deux de la fonction de densité (5.12) de la variable aléatoire $V(10000)$ (en milliers de π^3/s)

Intervalles	$(-\infty, 10]$	$(10, 15]$	$(15, \infty)$
Observations	12	10	8
Probabilités	0.483	0.326	0.191
Espérances	14.49	9.78	5.73

5.1.2 Distribution de Y si $\tau_{n+1} - \tau_n = 1$

Pour compléter cette section, on va essayer de déterminer la distribution de l'amplitude du signal Y , étant donné que $\tau_{n+1} - \tau_n = 1$. C'est-à-dire en supposant que le prochain

événement après l'instant $t = \tau_n$ aura lieu le jour suivant. Tel qu'il est mentionné ci-dessus, en pratique, seulement les observations quotidiennes du débit d'un cours d'eau sont généralement disponibles. Par conséquent, on veut déterminer la distribution de Y , étant donné qu'il y a (au moins) un événement du processus de Poisson dans l'intervalle $(\tau_n, \tau_{n+1}]$. Cette distribution est particulièrement importante lorsque la valeur de $X(\tau_n)$ est élevée et lorsqu'on veut estimer le risque d'inondation le lendemain.

Selon le processus de Poisson filtré et le processus de diffusion avec sauts, la probabilité qu'au moins un événement se produise en une journée est donnée par $1 - e^{-\lambda}$. Dans le cas du fleuve Delaware, cette probabilité est égale à environ 0,23, ce qui est relativement important. Cependant, bien que la distribution exponentielle n'a pas été rejetée pour la variable aléatoire T , on constate que parmi les 30 cas où la valeur de débit a augmenté, il y a 17 cas où cette augmentation a eu lieu un jour après τ_n . Par conséquent, on peut affirmer que la probabilité d'un tel événement est approximativement égale à $0,56 = (17/30)$, ce qui est beaucoup plus grand que 0,23. Ainsi, trouver la distribution de $Y|\{T = 1\}$ est d'une importance primordiale.

Un échantillon de taille 17 est plutôt petit pour fournir une estimation fiable des paramètres et trouver une distribution acceptable pour une variable aléatoire. Néanmoins, c'est toujours un exercice utile.

Basé sur cet échantillon de données, on constate que l'estimation ponctuelle du paramètre μ est maintenant $\hat{\mu} = 1/7796$. On note que cette valeur est beaucoup plus petite que l'estimation $\hat{\mu} = 1/5010$ obtenue précédemment. C'est que la taille du signal est en moyenne beaucoup plus grande lorsque ce signal se produit le lendemain.

On a effectué deux tests d'ajustement de khi-deux de l'hypothèse $Y \sim \text{Exp}(1/7796)$. Les résultats sont présentés dans les tableaux (5.7) et (5.8) en annexe A,.

La statistique du test est égale à 0,73 (respectivement 0,03, ce qui est remarquable) dans le cas du premier (respectivement deuxième) test. Les valeurs p correspondantes sont de 0,69 et 0,86. Ainsi, on peut conclure que le modèle exponentiel semble s'ajuster aux données très bien.

Cette distribution peut donc être utilisée pour estimer la probabilité que le débit augmentera encore de y pi³/s le lendemain du jour où il vient juste d'augmenter, pour tout $y \geq 0$. Cette estimation devrait s'avérer utile pour les gestionnaires de barrages, en particulier.

Tableau 5.7 Test d'ajustement du khi-deux de la distribution de la variable aléatoire Y (en milliers de π^3/s), étant donné que $T = 1$

Intervalles	$(0; 4]$	$(4; 8]$	$(8; 12]$	$(12; \infty)$
Observées	8	3	3	3
Probabilités	0,401	0,240	0,144	0,215
Espérées	6,82	4,08	2,45	3,65

Tableau 5.8 Second test d'ajustement du khi-deux de la distribution de la variable aléatoire Y (en milliers de π^3/s), étant donné que $T = 1$

Intervalles	$(0, 5]$	$(5, 10]$	$(10, \infty)$
Observations	8	4	5
Probabilités	0.473	0.249	2.77
Espérances	8.04	4.23	4.71

CHAPITRE 6 CONCLUSION

Dans la première partie de cette thèse, on a été en mesure d'améliorer de manière significative les prévisions hydrologiques à court terme produites par les processus de Poisson filtré en utilisant une distribution plus réaliste pour le temps d'inter-arrivée des signaux, en se servant donc des processus de renouvellement filtrés (Chapitre 3 (bien livrable 1)). Comme on l'a vu ci-dessus, le processus de Poisson filtré, en dépit de son manque de réalisme, est capable de donner des prévisions raisonnables à court terme. Une explication possible est le fait que les formules de prévision dérivées de ce modèle sont très faciles à mettre en œuvre. De plus, contrairement au processus de renouvellement filtré, il ne nécessite pas de décisions subjectives pour évaluer ses différents paramètres. Toutefois, le processus de Poisson filtré n'était pas en mesure de produire des prévisions aussi précises que celles qui découlent du processus de renouvellement filtré au cours de la période de pointe (crue) importante.

En plus d'être utilisés pour prévoir les valeurs de débit, les modèles considérés peuvent également nous donner une estimation de la probabilité que le débit d'un cours d'eau dépassera un certain seuil dans les prochains jours. Ce seuil peut être une valeur qui correspond à un débit au-delà duquel le risque d'inondation est très élevé.

Par ailleurs, on peut utiliser la fonction de réponse en (2.3) pour améliorer les prévisions. Pour ce faire, on doit être en mesure d'estimer le paramètre d dans cette fonction de réponse, puis obtenir les formules qui généralisent celles de (3.8) et (3.19), ce qui n'est probablement pas une tâche facile.

On a aussi considéré deux modèles stochastiques pour les variations des débits de cours d'eau. Les deux modèles supposent que les signaux conduisant à une augmentation du débit se produisent selon un processus de Poisson de taux λ , et que l'amplitude des signaux suit une loi exponentielle de paramètre μ . Toutefois, dans le cas du processus de Poisson filtré, on a utilisé une fonction de réponse exponentielle décroissante, alors qu'on a supposé que le débit évolue selon un processus de Wiener avec une dérive négative dans le second modèle considéré, soit le processus de diffusion avec sauts.

On s'est intéressé à obtenir une expression explicite de la probabilité que le débit dépasse un seuil donné le jour qui suit le débit courant $X(t) = x$ (Chapitre 4 (bien livrable 2)). En outre, parce que cette probabilité est particulièrement importante lorsque la valeur de débit courant est déjà assez grande, on a posé x égal à 10000 pi^3/s . Ensuite, on a calculé la probabilité que le débit à l'instant $t + 1$ augmentera d'au moins $y \text{ pi}^3/\text{s}$. Dans l'application sur le fleuve Delaware, on a choisi les valeurs $y = 5000, 10000$ et 20000 .

Après avoir estimé les différents paramètres des modèles, on a vu que les formules théoriques obtenues ont donné de très bons résultats par rapport aux fréquences observées correspondantes, ce qui n'était pas le cas lorsqu'on a utilisé la régression linéaire simple pour calculer la probabilité considérée.

Pour estimer les paramètres des modèles et calculer les fréquences observées, on a utilisé les données recueillies sur une période de 11 ans, ce qui est assez long. Cela était nécessaire parce qu'on avait besoin d'un nombre suffisant d'observations pour x qui appartiennent à l'intervalle $[10000, 11000)$. Pour obtenir peut-être même de meilleurs résultats, on pourrait étendre cette période. Cependant, en raison du changement climatique, en particulier, l'idée d'utiliser les données qui ont été recueillies, par exemple, il y a 50 ans, pourrait ne pas être utile.

Par la suite, on a étendu les résultats obtenus dans le chapitre 4 (bien livrable 2). En effet, tandis qu'on a obtenu les formules pour la probabilité qu'un débit donné augmentera d'au moins y unités en un jour, on a réussi à trouver des modèles acceptables pour la distribution de la valeur du débit au moment de la prochaine augmentation. Les résultats théoriques ont été appliqués avec succès au fleuve Delaware (Chapitre 5 (bien livrable 3)).

Les deux processus de Poisson filtré et de diffusion avec sauts ont été considérés comme des modèles pour les variations des débits. Les deux modèles s'avèrent appropriés. Les processus de diffusion avec sauts sont très populaires en mathématiques financières, mais n'ont pas été utilisés fréquemment en hydrologie. Avec la fonction de réponse qu'on a choisi pour le processus de Poisson filtré, le débit est censé diminuer de façon exponentielle et de façon déterministe lorsqu'aucun événement ne se produit pas dans un intervalle de temps donné. Cette hypothèse n'est, bien sûr, pas tout à fait réaliste. Dans le cas du processus de diffusion avec sauts, le débit diminue (en moyenne, car la dérive θ est négative) lorsqu'il n'y a pas de signaux, ce qui est logique, mais il le fait de manière stochastique. Ainsi, on peut affirmer que ce modèle est plus proche de la réalité.

D'autre part, en utilisant un processus de Wiener pour la partie continue du processus de diffusion avec sauts, cela implique que le débit pourrait, théoriquement, être négatif pour toute valeur de t , ce qui n'a pas de sens. Heureusement, en général la probabilité que le débit devienne négatif est négligeable. Toutefois, dans le cas de la variable aléatoire conditionnelle $X_D(\tau_{n+1})|X_D(\tau_n) = 10000$, on a vu que la probabilité d'observer une valeur négative du débit n'est pas négligeable. Pour éliminer ce problème, on pourrait utiliser un processus de diffusion qui ne peut pas devenir négatif, par exemple un mouvement brownien géométrique, qui est aussi un processus stochastique de base en mathématiques financières.

Enfin, on pourrait essayer d'obtenir une expression explicite de la fonction de densité de $X(\tau_{n+1})|X(\tau_n) = x$ pour d'autres valeurs de x et d'autres cours d'eau importants. Pour ce faire, on aurait besoin de plusieurs années d'observations à notre disposition.

Deuxième partie

ANALYSE DE SENSIBILITÉ ET D'INCERTITUDE DES MODULES DE MILIEUX HUMIDES D'UN MODÈLE HYDROLOGIQUE DISTRIBUÉ

CHAPITRE 7 INTRODUCTION

La performance d'un système d'ingénierie est souvent affectée par des incertitudes inévitables. Par conséquent, diverses méthodes ont été proposées pour quantifier les incertitudes du paramètre sur la sortie du modèle (Melching and Anmangandla (1992), Mailhot and Villeneuve (2003), Huang and Du (2008), Cacuci (2003) et Cacuci et al. (2005)). Selon Huang and Du (2008), l'efficience et la précision de ces méthodes d'analyse de l'incertitude est une question de compromis dans les applications d'ingénierie. C'est pourquoi, dans diverses disciplines de l'ingénierie, l'analyse d'incertitude est devenue une étape indispensable du développement de modèle car elle fournit un cadre pour évaluer les risques dans les modèles de prise de décision et de fiabilité (Morgan and Henrion (1990), Saltelli et al. (2004), Mailhot and Villeneuve (2003), Mahadevan (2000) et Putko et al. (2002)). Dans la modélisation hydrologique et de la qualité et de l'eau, l'incertitude est inhérente à la construction du modèle. Les phases générales de la construction du modèle sont les suivantes : (1) la planification du modèle, (2) la collecte de données et la conceptualisation du modèle, (3) la mise en place spécifique du modèle, (4) la calibration et la validation des paramètres du modèle, et (5) la simulation et l'évaluation (Peng and Zhao (2009)). Telles qu'elles sont identifiées par Beck (1987), les sources les plus importantes d'incertitude sont associées à la structure du modèle et les paramètres du modèle ; ces derniers étant la principale source selon Mailhot and Villeneuve (2003) et Melching and Anmangandla (1992). En tant que telle, cette forme d'incertitude suppose que la structure du modèle a été pratiquement identifiée. En effet, de nombreuses études ont porté sur l'évaluation de la propagation d'incertitude associée aux paramètres dans le modèle et la prédiction d'incertitude qui en résulte (Melching and Anmangandla (1992)). Deux principales méthodes ont été appliquées, la simulation de Monte-Carlo (MCS) qui nécessite beaucoup de temps de calcul et l'analyse d'incertitude issue de la méthode d'approximation par le développement en série de Taylor de premier degré. Cependant, cette dernière peut ne pas être la méthode préférable (ou la plus précise) pour les systèmes fortement non linéaires et sous des conditions extrêmes (Melching and Anmangandla (1992)).

À travers leurs fonctions de stockage d'eau, les milieux humides ont été reconnus comme unités de paysage potentiel visant à atténuer les inondations et les étiages sous l'évolution des conditions climatiques. Leurs impacts au niveau des bassins versants sont principalement motivés par le type et l'emplacement de la zone humide. Récemment, Fossey et al. (2015) et A. N. Rousseau (2013) ont mis en place et utilisé une plate-forme de modélisation hydrologique distribuée pour évaluer les fonctions hydrologiques des milieux humides isolés et riverains. La plate-forme de modélisation comprend des modules spécifiques de milieux hu-

mides qui sont intégrés dans HYDROTEL, un modèle hydrologique distribué (Bouda et al. (2013, 2012), Fortin et al. (2001a,b), Turcotte et al. (2007, 2003)), et PHYSITEL (Rousseau et al. (2011), Turcotte et al. (2001)), un système d'information géographique (GIS) pour soutenir la mise en œuvre des modèles distribués. Ce modèle divise la simulation des processus hydrologiques en six sous-modèles. Des informations détaillées sur la construction de cette plate-forme de modélisation peuvent être trouvées dans Fossey et al. (2015). Ils ont également présenté une analyse préliminaire de la sensibilité des paramètres des modules spécifiques aux milieux humides par rapport aux débits simulés à l'exutoire du bassin.

Cette étude s'appuie sur l'article de Fossey et al. (2015) en introduisant : (i) une analyse de sensibilité, induites par les variations paramétriques, par rapport aux débits simulés à différents segments ou tronçons du bassin versant de l'étude et (ii) une analyse d'incertitude attribuée aux paramètres basée sur des méthodes locales d'approximation par le développement en série de Taylor : la méthode d'approximation de premier ordre (*Mean-value, first-order, second-moment*-MFOSM) et la méthode d'approximation de second ordre (*mean-value, second-order, second-moment*-MSOSM) (Mailhot and Villeneuve, 2003). Les courbes des probabilités de dépassement (EP) sont générées pour des journées particulières, en utilisant les méthodes MFOSM et MSOSM. Ces courbes sont comparées à celles obtenues par simulation de Monte-Carlo (MCS). Ces deux approches méthodologiques, comme de nombreuses autres méthodes d'approximation, proviennent généralement du domaine de la fiabilité (Huang and Du, 2008), où elles sont utilisées pour estimer la probabilité d'occurrence des valeurs de réponse spécifiques compte tenu des fonctions de densité des paramètres.

Le chapitre 8 présente les méthodes existantes d'analyse de sensibilité et d'incertitude. On présente au chapitre 9 une brève description du modèle HYDROTEL et du bassin versant de l'étude ainsi qu'une discussion des résultats. Finalement, les principales conclusions de cette deuxième partie de la thèse sont présentées au chapitre 10.

CHAPITRE 8 L'ANALYSE D'INCERTITUDE ET DE SENSIBILITÉ

L'idée derrière l'analyse d'incertitude est de décrire tous les résultats possibles du modèle et leurs probabilités d'occurrence, tandis que celle de l'analyse de sensibilité est de déterminer, autour d'une région localisée de l'espace paramétrique, comment la réponse du modèle est sensible aux variations des valeurs des paramètres. En général, il est possible d'effectuer une analyse de sensibilité du modèle autour d'une solution en cours, et l'utiliser comme une partie d'une analyse d'incertitude de premier ordre (Loucks et al., 2005).

Les mesures de performance décrivant le débit de cours d'eau ou la qualité de l'eau varient dans le temps et l'espace comme une variable aléatoire dont on ne peut pas prédire la valeur avec certitude. Cependant, les distributions de probabilité des entrées ou paramètres du modèle et les réponses du modèle peuvent être déterminées dans une certaine plage de valeurs spécifiées. L'analyse d'incertitude consiste à identifier les caractéristiques des différentes distributions de probabilité et de toutes les sources d'incertitude qui contribuent à la distribution de probabilité conjointe de chaque variable d'entrée (paramètre) et de sortie du modèle.

On considère ici le cas où l'incertitude dans la réponse (la sortie du modèle) est due uniquement à l'incertitude dans les paramètres. Une analyse d'incertitude transmet un ensemble de valeurs d'entrée ou de paramètres choisis par l'intermédiaire d'un modèle afin d'obtenir des mesures statistiques de la distribution de la sortie qui en résulte. Cette distribution est utilisée pour estimer la probabilité que la sortie dépasse un seuil donné (Loucks et al., 2005; Da-Veiga, 2007). Le cadre probabiliste permet de décrire la distribution de probabilité de la sortie, les premiers moments, l'intervalle de confiance associé à un risque ou encore la probabilité de dépassement d'un seuil. Il est à noter que les méthodes peuvent varier en fonction de la forme attendue des résultats. Par conséquent, il est important d'identifier, dans un premier temps, l'objectif de l'analyse d'incertitude pour obtenir les résultats attendus sous une forme appropriée.

8.1 Les méthodes existantes de l'analyse d'incertitude

L'approche probabiliste repose sur la relation fonctionnelle déterministe suivante :

$$Y = g(\theta), \tag{8.1}$$

où le vecteur de sortie du modèle $Y = (Y_1, Y_2, \dots, Y_n)^T$ est entièrement déterminé par le vecteur des paramètres d'entrée du modèle $\theta = (\theta_1, \theta_2, \dots, \theta_m)^T$ qui contient les facteurs d'incertitude.

L'intérêt principal de l'analyse d'incertitude probabiliste est d'obtenir la distribution d'une fonction de réponse ou de performance Y , autrement dit, la probabilité de $Y \leq y$, compte tenu des distributions des paramètres aléatoires $\theta_1, \theta_2, \dots, \theta_m$. Puisque l'analyse d'incertitude probabiliste doit évaluer la relation fonctionnelle (8.1) un certain nombre de fois, le calcul intensif nécessaire est problématique.

La fonction de répartition cumulative de Y peut être calculée par une intégrale multidimensionnelle :

$$F_Y(y) = P[Y \leq y] = \int_{g(\theta) \leq y} f_{\theta}(s) ds, \quad (8.2)$$

où f_{θ} est la fonction de densité conjointe des paramètres aléatoires $\theta_1, \theta_2, \dots, \theta_m$, ces derniers sont supposés mutuellement indépendants.

En pratique, il est souvent difficile d'obtenir une solution analytique de l'équation (8.2). Cela est dû à la grande dimension et à la limite d'intégration $g(\theta) \leq y$. Pour cette raison, des simulations et des méthodes d'approximations sont utilisées. Ces dernières méthodes peuvent être classées en trois approches (Huang and Du, 2008) : (i) les méthodes basées sur l'échantillonnage (Dey and Mahadevan, 1998; McKay et al., 1979), telles que la simulation de Monte-Carlo (MCS), (ii) la méthode de correspondance des moments (Zhao and Ang, 2003; Wang et al., 2004) et (iii) la méthode basée sur le point le plus probable (*most-probable-point-based method*-MPP) (Hohenbichler et al., 1987; Du and Sudjianto, 2004).

La méthode de correspondance des moments approxime la distribution de Y en sélectionnant une distribution présupposée et en estimant les premiers moments. Parmi ces approches, les méthodes directes par intégration numérique, les méthodes d'estimation ponctuelle (Zhao and Ang, 2003; Wang et al., 2004) et les méthodes issues de développement en séries de Taylor (Zhao and Ono, 1999; Mailhot and Villeneuve, 2003; Huang and Du, 2008) ont été développées pour estimer les moments. Ce dernier type de méthodes est très efficace étant donné que la relation fonctionnelle est évaluée par le développement en série de Taylor de premier ou de second degré (Putko et al., 2002; Mailhot and Villeneuve, 2003).

La méthode MFOSM, qui utilise le développement en série de Taylor de premier degré au voisinage des valeurs moyennes des variables aléatoires, est l'approche la plus couramment utilisée; MFOSM est utilisée dans "*robust design optimization*" et l'analyse de fiabilité. La précision de la méthode de correspondance des moments, même avec une efficacité élevée, est généralement inférieure à celle des méthodes basées sur l'échantillonnage (Putko et al., 2002;

Mailhot and Villeneuve, 2003).

Entre-temps, les méthodes basées sur le point le plus probable (MPP) ont été développées pour obtenir un bon équilibre entre l'efficacité et la précision. Généralement utilisées dans "*reliability-based design optimization*" et "*structural reliability analysis*", cette méthode approxime la relation fonctionnelle du développement en série de Taylor au voisinage d'un point particulier (MPP), pour assurer le minimum de perte de précision. Diverses méthodes basées sur le MPP ont été proposées : la méthode de fiabilité du premier ordre (*the first order reliability method* - FORM), la méthode de fiabilité du second ordre (*the second order reliability method* - SORM) (Hohenbichler et al., 1987), et l'approximation récemment développée au voisinage du point de selle (*first-order Saddle point Approximation* - FOSA) (Du and Sudjianto, 2004). Le point MPP doit être localisé numériquement, ce qui fait que les méthodes basées sur ce point tendent à être coûteuses en calcul (Huang and Du, 2008).

Dans le cadre probabiliste, l'analyse d'incertitude est effectuée pour estimer la probabilité de dépassement (EP), P_E , le complément de la fonction de répartition dans l'équation (8.2). Elle est donnée par :

$$P_E(y) = P[Y > y] = \int_{g(\theta) > y} f_{\theta}(s) ds. \quad (8.3)$$

La probabilité P_E dans l'équation (8.3) peut être estimée en utilisant la simulation de Monte-Carlo (MCS), en générant un nombre de valeurs aléatoires du vecteur des paramètres θ , selon leurs fonctions de distribution de probabilité. La valeur de sortie Y est alors générée en fonction du type de modèle dans l'équation (8.1).

Ci-après, on présente brièvement deux méthodes existantes de correspondance des moments basées sur le développement en séries de Taylor de premier et second degré respectivement ; MFOSM et MSOSM. Ce sont parmi les méthodes les plus connues et utilisées en analyse d'incertitude des modèles hydrologiques et de la qualité d'eau.

8.1.1 La méthode de moments de premier ordre (MFOSM)

MFOSM est basée sur le développement en série de Taylor de premier degré de la relation fonctionnelle dans l'équation (8.1) au voisinage de la valeur moyenne (ou de référence) des paramètres d'entrée θ ; c'est-à-dire θ_0 .

L'équation (8.1) est alors approximée et linéarisée comme suit :

$$Y(\theta) \simeq Y(\theta_0) + \sum_{j=1}^m \left(\frac{\partial Y}{\partial \theta_j} \right)_{\vec{\theta}_0} (\theta_j - \theta_{j_0}), \quad (8.4)$$

En utilisant la relation fonctionnelle (aussi connue sous le nom de métamodèle) donnée dans l'équation (8.4), les probabilités de dépassement (EP) peuvent être calculées par MFOSM.

Les incertitudes sur les paramètres θ_j , $j = 1, \dots, m$, sont supposées suivre la distribution normale de paramètres $(\theta_{j0}, \sigma_{\theta_j}^2)$. À partir de l'équation (8.4), l'espérance et la variance de la relation fonctionnelle sont données par

$$E(Y) = \mu_Y \simeq Y(\theta_0), \quad (8.5)$$

$$Var(Y) = \sigma_Y^2 \simeq \sum_{j=1}^m B_j^2 var(\theta_j) + \sum_{j=1}^m \sum_{l=1}^m B_j B_l cov(\theta_j, \theta_l), \quad (8.6)$$

où

$$B_j = \left(\frac{\partial Y}{\partial \theta_j} \right)_{\vec{\theta}_0}. \quad (8.7)$$

Les paramètres sont supposés statiquement indépendants. Par conséquent, la variance dans l'équation (8.6) et l'écart-type deviennent

$$Var(Y) = \sigma_Y^2 \simeq \sum_{j=1}^m B_j^2 var(\theta_j), \quad (8.8)$$

$$\sigma_Y = \sqrt{\sum_{j=1}^m (B_j \sigma_{\theta_j})^2}. \quad (8.9)$$

Enfin, à partir de l'équation (8.4), la réponse Y est aussi normalement distribuée, puisqu'elle est une combinaison linéaire des θ_j . La probabilité de dépassement (EP) du modèle est

$$P_E(y) = P[Y > y] \simeq 1 - \phi\left(\frac{y - \mu_Y}{\sigma_Y}\right), \quad (8.10)$$

où $\phi()$ est l'intégrale d'une fonction normale standard.

On peut alors calculer les probabilités de dépassement à l'aide de l'équation (8.10), en utilisant les valeurs obtenues de la moyenne μ_Y (8.5) et de l'écart-type σ_Y (8.9) et en variant la valeur de y .

Il existe une méthode de moments de premier ordre (*advanced first-order second-moment* (AFOSM)) qui a été proposée (voir Rackwitz, 1976; Melching and Anmangandla, 1992; Mailhot and Villeneuve, 2003). La première idée de cette méthode d'AFSOM a été suggérée par Hasofer and Lind (1974). Cette idée repose sur la linéarisation de la relation fonctionnelle en

utilisant le développement en série de Taylor à un point sélectionné sur la surface de rupture. La surface de rupture est la surface qui délimite la région de l'espace paramétrique θ où $Y(\theta_j) > y$ et $Y(\theta_j) < y$, pour $j = 1, \dots, m$. Ce point correspond à la distance la plus courte dans l'espace normalisé entre le point de la valeur moyenne et le point de rupture le plus probable de la surface de rupture. La détermination de ce point est la principale difficulté dans la méthode AFOSM (Mailhot and Villeneuve, 2003).

8.1.2 La méthode de moments de second ordre (MFOSM)

MSOSM est basée sur le développement en série de Taylor de second degré de la relation fonctionnelle (8.1) au point de la valeur de référence θ_0 dans l'espace paramétrique. La relation fonctionnelle $g(\theta)$ (le métamodèle d'ordre 2) est donnée par

$$\begin{aligned} Y(\theta) &\simeq Y(\theta_0) + \sum_{j=1}^m \left(\frac{\partial Y}{\partial \theta_j} \right)_{\vec{\theta}_0} (\theta_j - \theta_{j_0}) \\ &+ \frac{1}{2} \sum_{j=1}^m \sum_{l=1}^m \left(\frac{\partial^2 Y}{\partial \theta_j \partial \theta_l} \right)_{\vec{\theta}_0} (\theta_j - \theta_{j_0})(\theta_l - \theta_{l_0}). \end{aligned} \quad (8.11)$$

Le développement suivant de la méthode est donné par Mailhot and Villeneuve (2003). On définit les variables suivantes :

$$\Delta Y = Y(\theta) - Y(\theta_0), \quad (8.12)$$

$$u_j = \frac{\theta_j - \theta_0}{\sigma_j}, \quad (8.13)$$

$$a_j = \sigma_j B_j, \quad (8.14)$$

$$a_{jl} = \frac{1}{2} \sigma_j \sigma_l B_{jl}, \quad (8.15)$$

où

$$B_{jl} = \left(\frac{\partial^2 Y}{\partial \theta_j \partial \theta_l} \right)_{\vec{\theta}_0}. \quad (8.16)$$

L'équation (8.11) peut être écrite en fonction de la distribution normale standard u_j comme suit :

$$\Delta Y \simeq \sum_{j=1}^m a_j u_j + \sum_{j=1}^m \sum_{l=1}^m a_{jl} u_j u_l, \quad (8.17)$$

ou exprimée sous forme matricielle :

$$\Delta Y = a^T u + u^T A u, \quad (8.18)$$

avec $a = (a_1 \dots a_m)^T$, $u = (u_1 \dots u_m)^T$ et

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mm} \end{pmatrix}.$$

La matrice A est supposée diagonalisable, soit $Q^T A Q = \Lambda$, où Q est une matrice formée des vecteurs propres q_j de la matrice A mis en colonnes et Λ est une matrice diagonale avec les valeurs propres λ_j de A . L'équation (8.18) peut alors être exprimée par

$$\Delta Y \simeq \sum_{j=1}^m a^T q_j x_j + \sum_{j=1}^m \lambda_j x_j^2, \quad (8.19)$$

où $x_j = q_j^T u$.

Cette équation définit la surface de rupture, pour ΔY donnée dans l'espace jusqu'au deuxième ordre en u . Considérant que x_j est décrite par une distribution normale standard, la probabilité que $\Delta Y > \delta$ est donnée par

$$P[\Delta Y > \delta] = \int_{\Omega(\theta)} f_u(t) dt, \quad (8.20)$$

où f_u est une fonction standard de probabilité multinormale centrée à zéro et $\Omega(\theta)$ est le domaine d'intégration correspondant à l'espace paramétrique où $\Delta Y > \delta y$.

Finalement, soit

$$W = \sum_{j=1}^m \lambda_j (x_j - x_{j0})^2 = \Delta Y + \sum_{j=1}^m \lambda_j x_{j0}^2, \quad (8.21)$$

avec $x_{j0} = -\frac{a^T q_j}{2\lambda_j}$.

La forme quadratique W dans (8.21) représente une surface centrée aux points x_{j0} , $j = 1, \dots, m$ dans un espace paramétrique à m dimensions et délimite les régions où $W < w$ ou $W > w$.

Imhof (1961) a développé une méthode pour évaluer la fonction de distribution cumulative des formes quadratiques des variables normales, telle que W . Basée sur cette méthode, la probabilité de dépassement, définie par l'équation (8.20) est alors analogue à la probabilité $P(W > w)$ qui est estimée par

$$P(W > w) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \frac{\sin \varphi(v)}{v \rho(v)} dv, \quad (8.22)$$

où

$$\varphi(v) = \frac{1}{2} \sum_{i=1}^m \left[\arctan(\lambda_i v) + \frac{x_{i0}^2 \lambda_i v}{(1 + \lambda_i^2 v^2)} - \frac{1}{2} w v \right], \quad (8.23)$$

$$\rho(v) = \exp \left[\frac{1}{2} \sum_{k=1}^m \frac{(x_{k0} \lambda_k v)^2}{(1 + \lambda_k^2 v^2)} \right] \prod_{i=1}^m (1 + \lambda_i^2 v^2)^{1/4}. \quad (8.24)$$

Si le seuil w spécifié dans l'espace propre et orthogonal est donné par :

$$w = \delta + \sum_{j=1}^m \lambda_j x_{j0}^2, \quad (8.25)$$

Alors, la probabilité $P(W > w)$ sera transformée en $P(Y(\theta) > Y(\theta_0) + \delta)$. La fonction de probabilité de dépassement (EP) de la variable Y est alors estimée par l'expression dans l'équation (8.22), en utilisant l'intégration numérique et en faisant varier les valeurs de δ .

8.2 L'analyse de sensibilité

Une analyse de sensibilité décrit combien les valeurs de la réponse du modèle sont affectées par les variations dans les valeurs d'entrée. Diverses approches ont été développées ; la plupart d'entre elles évaluent l'effet des variations dans un seul paramètre ou la valeur de la variable d'entrée en supposant qu'il n'y a aucun changement dans toutes les autres variables d'entrées. Si la plage d'incertitude d'intérêt est reliée seulement à certaines données de sortie, alors seulement les données d'entrée qui influencent significativement ces données de sortie doivent être incluses dans l'analyse de sensibilité. Le coefficient ou l'indice de sensibilité est une mesure de la sensibilité du modèle. Les analyses de sensibilité les plus simples, généralement, examinent les erreurs dans les paramètres un à la fois ou en groupes (Loucks et al., 2005).

Les méthodes d'analyse de sensibilité sont généralement regroupées (Saltelli et al., 2000, 2004) en méthodes globales ou locales. Dans la modélisation hydrologique et de la qualité de l'eau, l'accent est mis sur le classement de l'effet des incertitudes associées aux paramètres sur la variation de la réponse du modèle. Ainsi, il est nécessaire de trouver les indices de sensibilité appropriés qui sont dérivés à partir du modèle et dont l'interprétation apporte une réponse à ce problème.

On décrit dans ce qui suit la méthode locale de premier ordre de l'analyse de sensibilité.

8.2.1 L'analyse de sensibilité du premier ordre

Les méthodes locales sont basées soit sur l'analyse différentielle ou les méthodes de rang nominal. Dans l'analyse différentielle, les indices de sensibilité sont considérées comme les valeurs numériques des dérivées partielles approximatives des réponses du modèle par rapport aux perturbations imposées dans les paramètres du modèle. Alors que les indices de sensibilité des méthodes de rang nominal sont décrites comme étant le rapport des variations de la réponse du modèle et les variations des paramètres, étant donné leur valeur de référence (Bouda et al., 2013).

La méthode locale propose d'estimer le coefficient de sensibilité (Saltelli et al. (2000)) par la dérivée :

$$\left(\frac{\partial Y}{\partial \theta_j} \right)_{\vec{\theta}_0}, \quad j = 1, \dots, m, \quad (8.26)$$

où Y est la relation fonctionnelle approximée par le développement en série de Taylor de premier degré (8.4). Cette mesure estime la sensibilité de la réponse du modèle Y à la variation de θ_j .

Cette mesure peut être normalisée comme suit :

$$\frac{\theta_{j0}}{Y(\theta_0)} \left(\frac{\partial Y}{\partial \theta_j} \right)_{\vec{\theta}_0}, \quad j = 1, \dots, m, \quad (8.27)$$

où θ_{j0} est la j^e valeur de θ_0 .

En utilisant le cadre probabiliste pour examiner la sensibilité, il est possible de prendre en compte l'écart-type du paramètre θ_j et de la réponse Y . Le coefficient de sensibilité (Da-Veiga, 2007) devient

$$\frac{\sigma(\theta_j)}{\sigma(Y)} \left(\frac{\partial Y}{\partial \theta_j} \right)_{\vec{\theta}_0}, \quad j = 1, \dots, m. \quad (8.28)$$

Dans cette étude, on définit la fonction de sensibilité $s_{i,j}$ de la réponse Y correspondant à une variation $\Delta\theta$ dans la valeur du paramètre θ par rapport à sa valeur de référence θ_0 (voir Bouda et al., 2013) par :

$$s_{i,j} = \frac{\partial Y_i / Y_i(\theta_{j0})}{\partial \theta_j / \theta_{j0}} = \frac{\partial Y_i}{\partial \theta_j} \frac{\theta_{j0}}{Y_i(\theta_{j0})}, \quad i = 1, \dots, n, j = 1, \dots, m. \quad (8.29)$$

où Y_i représente la valeur de la i -ème journée (en considérant des observations quotidiennes).

En pratique, la fonction de sensibilité peut être simplifiée comme suit :

$$s_{i,j} = \frac{\theta_{j0}}{2\Delta\theta_j Y_i(\theta_{j0})} (Y_i(\theta_{j0} + \Delta\theta_j) - Y_i(\theta_{j0} - \Delta\theta_j)). \quad (8.30)$$

Ainsi, pour mesurer l'importance de l'effet de chaque paramètre sur la sensibilité du modèle et ainsi identifier le paramètre le plus influent (important) sur la réponse du modèle, l'indice de sensibilité $S_j(\theta)$ est calculé à partir des équations (8.29), (8.30) comme suit :

$$S_j(\theta) = \sqrt{\frac{1}{n} \sum_{i=1}^n s_{i,j}^2}. \quad (8.31)$$

$S_j(\theta)$ prend ses valeurs entre 0 et 1 ; plus la valeur de l'indice augmente et se rapproche de 1, l'impact de l'incertitude du paramètre sur la sortie du modèle est important.

CHAPITRE 9 ANALYSE D'INCERTITUDE ET DE SENSIBILITÉ DES MODULES DE MILIEUX HUMIDES DU MODÈLE HYDROTEL

9.1 Le modèle hydrologique HYDROTEL

HYDROTEL est un modèle hydrologique distribué de bassins versants qui est actuellement utilisé pour les apports et les prévisions hydrologiques par Hydro-Québec, la principale entreprise publique d'électricité du Québec, et le centre d'expertise hydrique du Québec (CEHQ) (Fortin et al., 2001a; Turcotte et al., 2003, 2007; Bouda et al., 2012). La construction de la base de données nécessite des données hydro-météorologiques et physiographiques limitées, et est effectuée par un outil basé sur le système d'information géographique (SIG) semi-automatique appelé PHYSITEL (Turcotte et al., 2001; Rousseau et al., 2011). Ce dernier est utilisé pour discrétiser un bassin versant en unités hydrologiques relativement homogènes (UHRH) et des segments de rivières interconnectés (SR), formant les unités de calcul spatiales (connues aussi sous le nom d'objets hydrologiques). Ces unités et les segments sont ensuite paramétrés pour HYDROTEL. Chacune des UHRH est caractérisée par son sol dominant et les pourcentages de chaque classe de couverture terrestre (Bouda et al., 2012; A. N. Rousseau, 2013). Le modèle fonctionne sur une échelle de temps quotidienne ou une échelle de trois heures. La précipitation quotidienne, les températures maximales et minimales et les débits journaliers à l'exutoire du bassin versant sont les variables principales introduites dans la base de données hydro-météorologiques (Fossey et al.) (2015) et (Fortin et al., 2001a,b).

En pratique, le modèle est basé sur des algorithmes développés à partir des approximations conceptuelles ou empiriques des processus physiques. HYDROTEL comprend six sous-modèles computationnels de base pour simuler un processus hydrologique spécifique : (1) l'interpolation des données météorologiques, (2) l'évolution de la couverture de neige, (3) l'évapotranspiration potentielle, (4) la bilan de l'eau vertical (la teneur en eau du sol; *soil column moisture*), (5) l'écoulement de surface, et (6) l'acheminement de débit à travers le réseau de la rivière (Fortin et al., 2001a).

Fossey et al. (2015) montrent que l'ajout des deux modules de milieux humides MH, à savoir, les modules de MH isolés et MH riverains augmente la capacité du modèle à reproduire les caractéristiques de l'hydrogramme et performe la plate-forme de la modélisation tout en évaluant l'impact hydrologique de ces MH. Basé sur le concept des milieux hydrologiques humides équivalents, ces modules opèrent au niveau des UHRH. Le module des MH riverains est partiellement intégré dans le bilan hydrique d'une UHRH tout en étant connecté directement au modèle des débits de la rivière. Le module des MH isolés est défini comme un volume

de contrôle dans le bilan hydrique d'une UHRH tout en étant lié à l'écoulement de surface (Fossey et al.) (2015). Toutes les caractéristiques des modules des MH et de leur intégration au sein de la plate-forme de modélisation, ainsi que l'évaluation du modèle (calibration et validation) est également décrite par (Fossey et al.) (2015).

9.2 Le bassin versant à l'étude

Le bassin versant de la rivière Bécancour couvre une superficie de 2 596 km² et est situé dans les basses-terres de Saint-Laurent au sud du Québec. La zone humide est de 300 km², soit environ 12 % de la superficie totale du bassin versant. Trois segments d'intérêt sur le bassin versant ont été définis pour le besoin des procédures de paramétrage, calibration et de validation : les sous-bassins versants commençant de l'amont à l'aval, ou les tronçons (TR255, TR102 et TR70), en plus de l'exutoire (TR1) ; voir la figure (9.1) en annexe A,. Dix-huit paramètres ont été calibrés, quinze affectant le volume des eaux de ruissellement et trois associés à l'acheminement de débit.

Les données disponibles utilisées pour évaluer le modèle HYDROTEL dans le cas des tronçons TR70, TR102 et TR255 s'étalent sur une période de 30 ans (1970-2000), 9 ans (2000-2009) et 39 ans (1970-2009), respectivement. Les données observées de l'ensemble du bassin versant comprenaient les débits journaliers et les séries météorologiques tout au long de la période (1970/01/01-2010/09/30), donc 40 ans de données. Pour plus de détails sur la calibration et la validation du modèle HYDROTEL pour la rivière Bécancour, voir Fossey et al. (2015).

Notre objectif étant la réalisation d'une analyse de sensibilité et d'incertitudes d'un nouveau module spécifique aux milieux humides nouvellement intégrés dans le modèle hydrologique HYDROTEL. Les méthodes employées pour effectuer l'analyse d'incertitude à l'aide de la simulation de Monte-Carlo (MCS) et les deux méthodes d'approximation par le développement en série de Taylor de premier ou second ordre (MFSOM et MSOSM), ainsi que l'analyse de sensibilité seront présentées dans la section suivante.

9.3 Analyse d'incertitude

L'analyse d'incertitude permet d'estimer l'incertitude du débit simulé autour de sa valeur de référence comme une résultante des incertitudes attribuées aux paramètres. Il est important d'effectuer cette analyse pour pouvoir déceler un effet significatif sur le débit simulé dû à des changements spécifiques sur le bassin versant.

L'ensemble des paramètres est déjà identifié dans cette étude et il y en a dix : cinq pour les

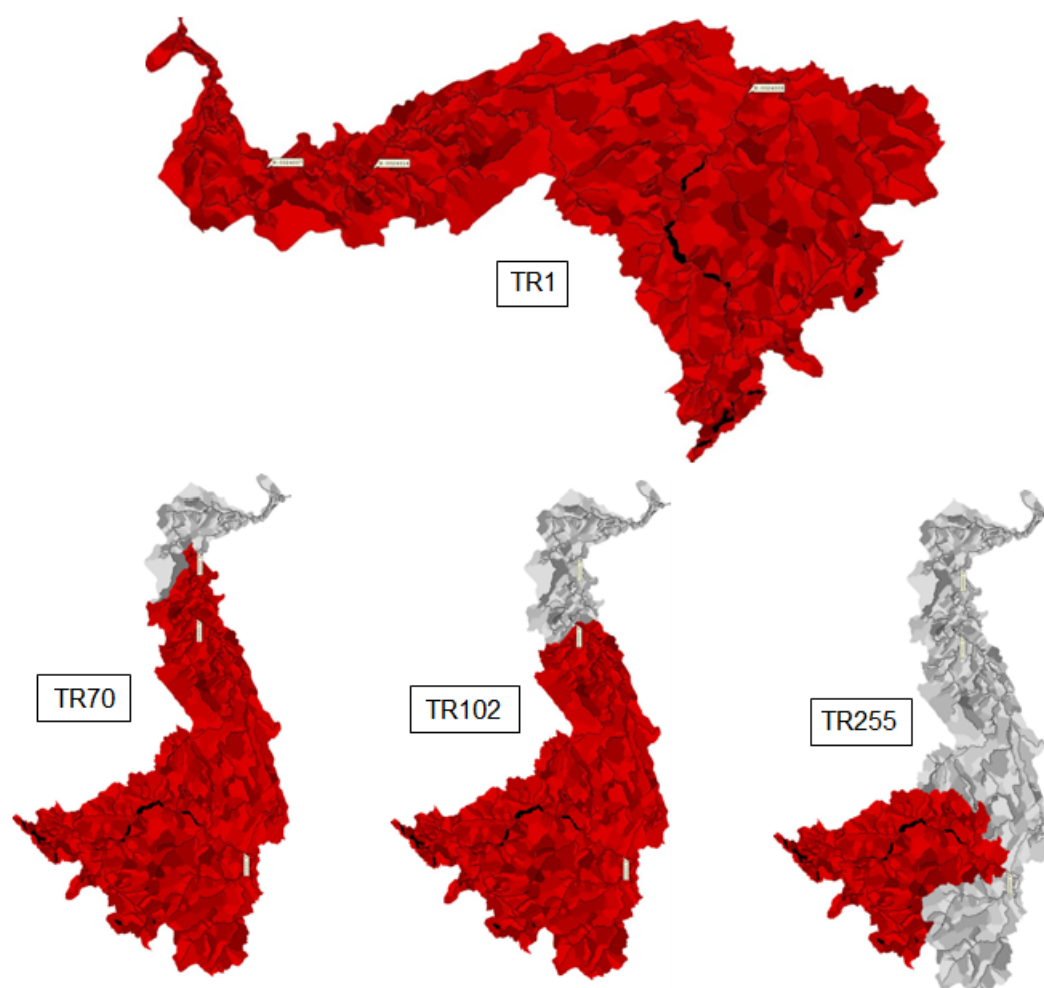


Figure 9.1 Le bassin versant TR1 de la rivière Bécancour et ses sous-bassins ou tronçons TR70, TR102 et TR255

MH isolés et cinq pour les MH riverains. La description de ces paramètres θ est présentée dans le tableau (9.1) avec leurs valeurs de référence et incertitudes.

La simulation des débits de référence $Y(\theta_0)$ des deux milieux humides est réalisée, sur une période de 40 ans (1970/01-2010/09), par le modèle HYROTEL en considérant les valeurs de référence des paramètres θ_j .

La variation du débit simulé serait induite à partir de différentes valeurs d'incertitudes paramétriques $\Delta\theta$ choisies arbitrairement : $(\pm 0,0625)$, $(\pm 0,125)$, $(\pm 0,25)$ et $(\pm 0,50)$. Chaque incertitude est le double de la précédente. Ainsi, en variant positivement et négativement la valeur de chaque paramètre, le débit résultant serait simulé par le modèle tel $Y(\theta_j) = Y(\theta_j + \Delta\theta_j)$ ou $Y(\theta_j) = Y(\theta_j - \Delta\theta_j)$.

Les résultats préliminaires ont montré que l'incertitude de $\pm 50\%$ donne les plus grandes variations des débits simulés comparativement aux autres incertitudes. En fait, les résultats de simulation obtenus pour l'ensemble du bassin versant, en considérant chacune des incertitudes révèlent que la variance des débits simulés est légèrement différentes et que la réponse du modèle est tout à fait linéaires aux changements des paramètres.

Par conséquent, pour alléger la présentation des résultats, la variation des débits simulés résultant d'une incertitude de $\pm 50\%$ sur les paramètres par rapport à leur valeur de référence sera considérée. Cette incertitude ($\pm 50\%$) paramétrique donne une variation significative du débit simulé.

L'importance de l'incertitude propagée à travers le modèle sur les débits simulés est effectuée à l'aide des méthodes MCS, MFOSM et MSOSM. Ces méthodes permettront d'estimer la fonction de distribution et la probabilité de dépassement des débits simulés en divers points correspondant aux tronçons du bassin versant et de quantifier l'erreur sur la réponse du modèle, cette dernière serait d'autant plus importante s'il est possible d'évaluer des différences significatives des variations des débits simulés.

9.3.1 La méthode de Monte-Carlo

En introduisant les incertitudes sur tous les paramètres, un échantillon aléatoire d'un ensemble de paramètres d'entrée est tiré à partir de la distribution conjointe supposée normale, et l'échantillon correspondant de la réponse du modèle (les débits) est simulé à partir du métamodèle (8.11). L'idée est de simuler le débit pour différents jeux de paramètres un nombre de fois suffisant (soit par exemple 10^5 simulations), puis obtenir les probabilités de dépassement d'un seuil de débit pour tous les jeux de paramètres considérés. Les distributions normales tronquées ont été utilisées pour les différentes valeurs des paramètres afin de ne

tenir compte que de leur valeurs admissibles.

Tableau 9.1 Description des paramètres θ_j , leur valeurs de référence θ_{j0} et leurs incertitudes $\Delta\theta_j$ des MH isolés et riverains du bassin versant de la Bécancour

MH	Paramètres θ_j	Description	Valeurs de référence θ_{j0}	Variation $\Delta\theta_j$
Isolé	C-EV	Coefficient pour l'ETP (évapotranspiration potentielle)	0.6	0.3
	C-PROD	Coefficient du MH	0.1	0.05
	FRAC	Coefficient de relation entre surface maximale et surface normale	0.3	0.15
	KSAT-BS	Conductivité à saturation de la base du MH (mm/h)	0.5	0.25
	RAV	Coefficient de relation entre surface et volume	0.96535	0.48267
Riverain	FRAC	Coefficient de relation entre surface maximale et surface normale	0.3	0.15
	KSAT-BK	Conductivité à saturation de la berge (mm/h)	25	15.5
	KSAT-BS	Conductivité à saturation de la base du MH (mm/h)	0.5	0.25
	WETDMAX	Hauteur d'eau maximale (m)	0.5	0.25
	WETDNOR	Hauteur d'eau normale (m)	0.3	0.05

9.3.2 La méthode MFOSM

La réalisation de la méthode d'approximation d'ordre un se fera par les étapes suivantes :

1. La simulation des débits en trois points du bassin versant délimitant les trois tronçons (TR72, TR102 et TR255) ainsi qu'à l'exutoire (TR1), à partir du métamodèle d'ordre un (8.4), au moyen des valeurs de référence des paramètres $Y(\theta_0)$ et les valeurs variées de paramètres, positivement $Y(\theta_j^+) = Y(\theta_j + \Delta\theta_j)$ et négativement $Y(\theta_j^-) = Y(\theta_j - \Delta\theta_j)$.
2. Calcul des dérivées premières :

$$\left(\frac{\partial Y}{\partial \theta_j}\right)_{\vec{\theta}_0} = \frac{Y(\theta_{j0} + \Delta\theta_j) - Y(\theta_{j0} - \Delta\theta_j)}{2\Delta\theta_j}. \quad (9.1)$$

Tableau 9.2 Les surfaces des MH drainées et occupées par les secteurs considérés du BV de la rivière de Bécancour

	BV-exutoire	TR1	TR70	TR102	TR255
MH (km^2)	2579		2306	2164	918
MH (%)	38		40	38	29
MH isolés (%)	24		25	23	14
MH riverains (%)	14		15	15	15

3. Introduction d'un seuil y correspondant aux variations de débit par rapport à sa valeur de référence spécifique à un jour particulier.
4. Calcul de la probabilité de dépassement pour le débit simulé aux quatre tronçons considérés à l'aide de l'équation (8.4).
5. Répéter les étapes (3) et (4) en variant la valeur du seuil y .

9.3.3 La méthode MSOSM

Les étapes nécessaires à l'application de la méthode d'approximation d'ordre deux se résument dans ce qui suit :

1. La simulation des débits aux quatre tronçons (TR72, TR102 et TR255) et l'exutoire (TR1), à partir du métamodèle d'ordre deux (8.11), au moyen des valeurs de référence des paramètres $Y(\theta_0)$ et les valeurs variées de paramètres, positivement $Y(\theta_j^+)$ et négativement $Y(\theta_j^-)$.
2. Calcul des dérivées premières (selon l'équation (9.1)), secondes et mixtes autour du paramètre de référence θ_0 :

$$\left(\frac{\partial^2 Y}{(\partial \theta_j)^2} \right)_{\vec{\theta}_0} = \frac{Y(\theta_{j0} + \Delta \theta_j) - 2Y(\theta_{j0}) + Y(\theta_{j0} - \Delta \theta_j)}{(\Delta \theta_j)^2}. \quad (9.2)$$

$$\left(\frac{\partial^2 Y}{\partial \theta_j \partial \theta_l} \right)_{\vec{\theta}_0} = \frac{1}{4\Delta \theta_j \Delta \theta_l} \begin{pmatrix} Y(\theta_{j0} + \Delta \theta_j, \theta_{l0} + \Delta \theta_l) + Y(\theta_{j0} - \Delta \theta_j, \theta_{l0} - \Delta \theta_l) \\ -Y(\theta_{j0} + \Delta \theta_l, \theta_{l0} - \Delta \theta_l) - Y(\theta_{l0} - \Delta \theta_j, \theta_{l0} + \Delta \theta_j) \end{pmatrix} \quad (9.3)$$

3. Calcul des vecteurs et valeurs propres de la matrice A et initialisation de la valeur δ à zéro.
4. Calcul de w avec l'équation (8.25) et la probabilité de dépassement par les équations (8.22), (8.23) et (8.24).
5. Répétition de l'étape précédente en variant la valeur de δ .

9.3.4 Résultats et discussion

L'analyse d'incertitude est effectuée pour estimer la probabilité que le débit simulé dépasse un seuil donnée lorsque les incertitudes sur les paramètres sont connues. Les méthodes MFOSM et MSOSM sont appliquées pour calculer les probabilités de dépassement (EP). La méthode des simulations Monte-Carlo (MCS) est utilisée pour valider les résultats obtenus par MFOSM et MSOSM.

La comparaison entre les méthodes MCS, MFOSM et MSOSM est visualisée par les courbes des fonctions (EP) des débits journaliers. Ces probabilités de dépassement sont calculées pour deux jours particuliers : le jour du débit maximum et le jour du débit minimum. Les figures (9.2) et (9.3) illustrent les courbes de ces probabilités (EP) pour ces jours particuliers, respectivement, calculées pour le bassin versant à l'exutoire TR1. Les résultats obtenus pour les tronçons TR70, TR102 et TR255 sont présentés par les figures ((9.4), (9.5)), ((9.6), (9.7)) et ((9.8), (9.9)) respectivement.

Les deux journées particulières sont choisies de façon à correspondre aux phénomènes de crue et d'étiage. Le débit maximum et minimum étant déterminés sur une période de 40 ans.

Les méthodes MCS et MSOSM donnent en général, des résultats similaires des probabilités de dépassement (EP). Les résultats de la méthode du MFOSM parfois sous-estiment légèrement les valeurs de ces probabilités (EP) (Figures (9.4), (9.5), (9.6) et (9.7)) et parfois elle les surestiment (Figure (9.3)).

Il est évident que la méthode de MSOSM conduira à une meilleure estimation de la courbe des probabilités de dépassement (EP). Puisque les approximations utilisées dans les méthodes MFOSM et MSOSM sont basées sur le développement en série de Taylor de la fonction de réponse au voisinage de la valeur moyenne ou de référence des paramètres, la méthode MSOSM, se basant sur une approximation quadratique, utilise plus d'informations que la méthode MFOSM, la surface de rupture sera alors mieux définie comparativement à une approximation linéaire.

Il faut noter que les trois méthodes MFOSM, MSOSM et MCS conduisent à des probabilités de dépassement presque identiques obtenue dans le cas du tronçon TR255 (Figure (9.8) et (9.9)). Il est très probable que les milieux humides isolées sont principalement responsables du comportement non linéaire pour ce sous-bassin hydrographique. En fait, la surface de drainage des deux types de milieux humides est similaire pour le tronçon TR255. Pour tous les autres (W1, SW70, SW102), la surface des milieux humides isolées représentent deux fois celle des milieux humides riverains.

L'analyse de sensibilité complétera l'information recueillie par l'analyse d'incertitude en étu-

diant comment la variation du débit simulé peut résulter de différentes sources d'incertitude et combien le modèle serait sensible à ces incertitudes accompagnant les paramètres d'entrée.

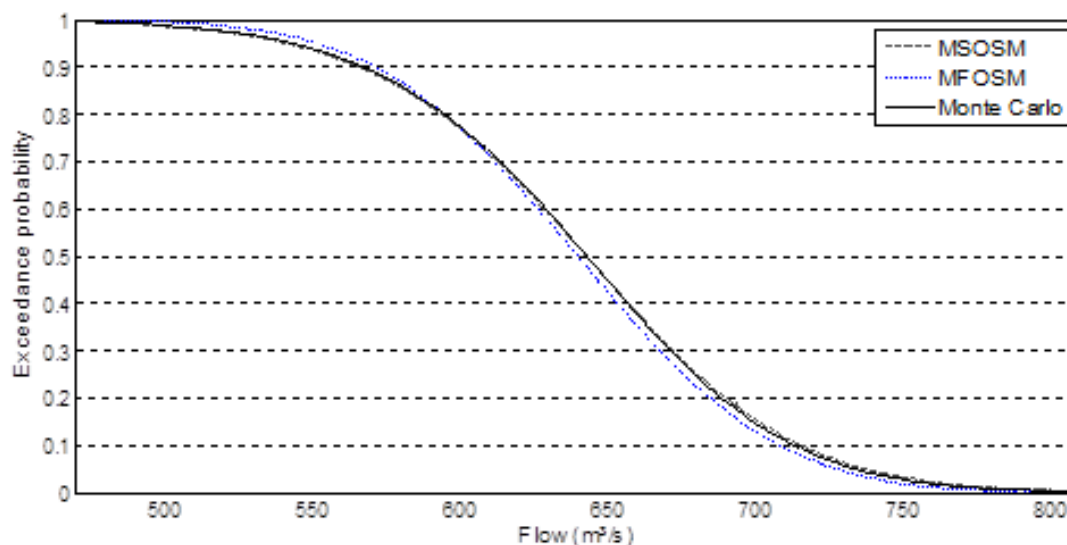


Figure 9.2 Comparaison des valeurs des EP sur le TR1 pour la journée du débit maximum résultant des différentes méthodes d'analyse d'incertitude.

9.4 Analyse de sensibilité

Les résultats de l'analyse de sensibilité sont présentés pour évaluer l'impact des incertitudes associées à chaque paramètre sur la réponse du modèle ou le débit simulé par HYDROTEL et par conséquent, identifier les paramètres les plus influents à la variation du débit simulé. Cette information est utile pour pouvoir procéder au choix final des paramètres à considérer lors du calage du modèle.

L'approche de l'analyse de sensibilité du premier ordre (voir Bouda et al., 2013) est appliquée sur les milieux humides de l'ensemble du bassin versant de la rivière Bécancour (l'exutoire TR1) et réalisée sur trois tronçons (TR70, TR102 et TR255).

La valeur d'un seul paramètre à la fois a été modifiée, l'autre étant fixé à la valeur attribuée (valeurs de référence). Toutes les données sont normalisées par le pourcentage des régions occupées ou drainées par les milieux humides isolés et riverains, pour tous les tronçons considérés (voir tableau 9.2 en annexe A). Enfin, les paramètres les plus influents sont déterminés.

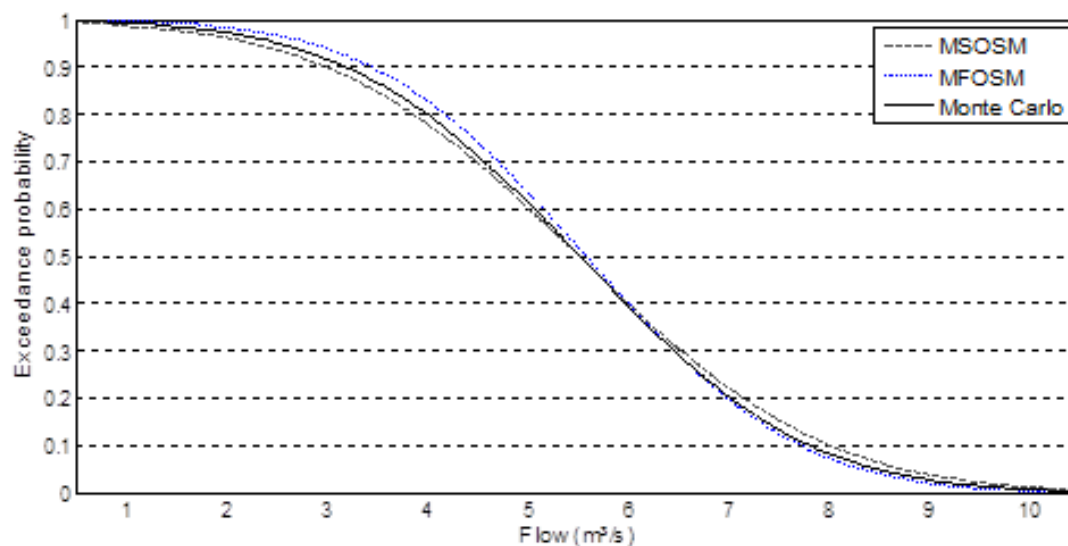


Figure 9.3 Comparaison des valeurs des EP sur le TR1 pour la journée du débit minimum résultant des différentes méthodes d'analyse d'incertitude.

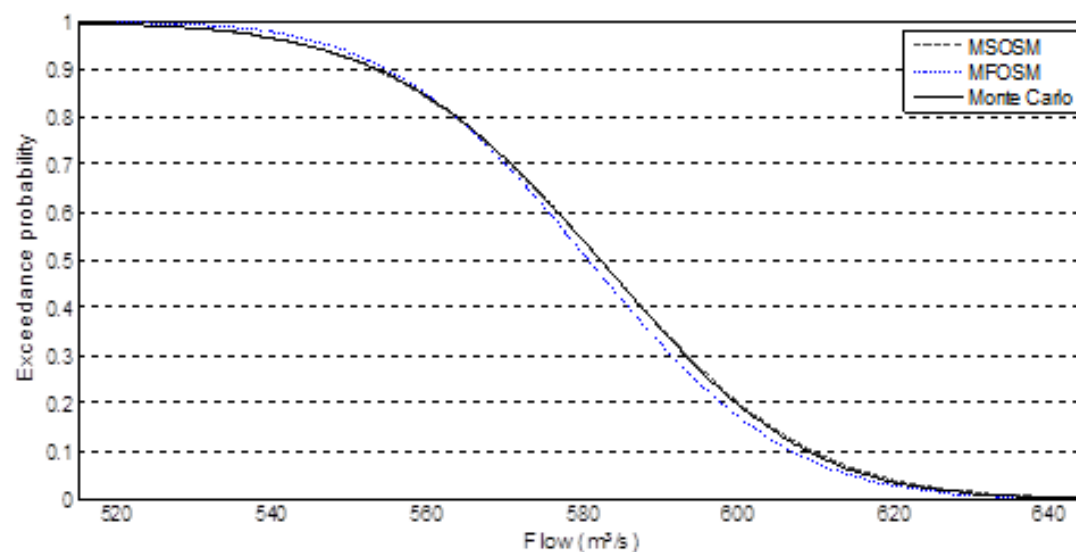


Figure 9.4 Comparaison des valeurs des EP sur le TR70 pour la journée du débit maximum résultant des différentes méthodes d'analyse d'incertitude.

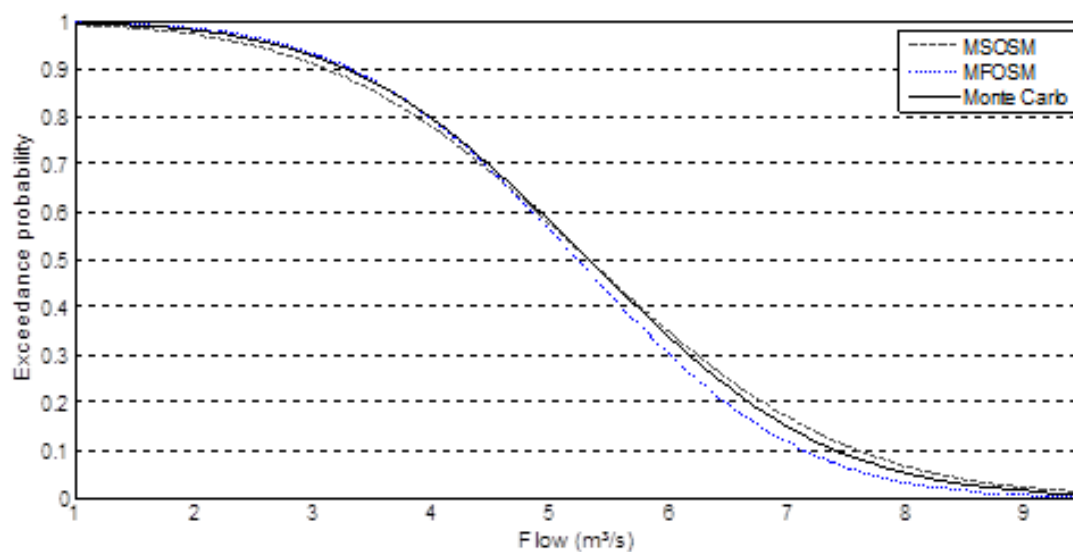


Figure 9.5 Comparaison des valeurs des EP sur le TR70 pour la journée du débit minimum résultant des différentes méthodes d'analyse d'incertitude.

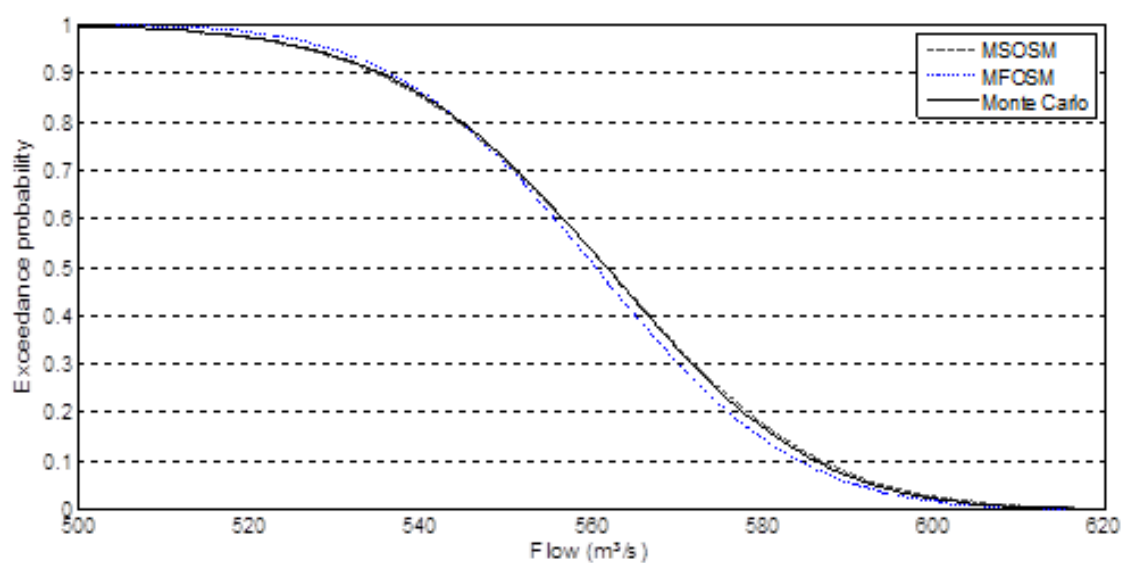


Figure 9.6 Comparaison des valeurs des EP sur le TR102 pour la journée du débit maximum résultant des différentes méthodes d'analyse d'incertitude.

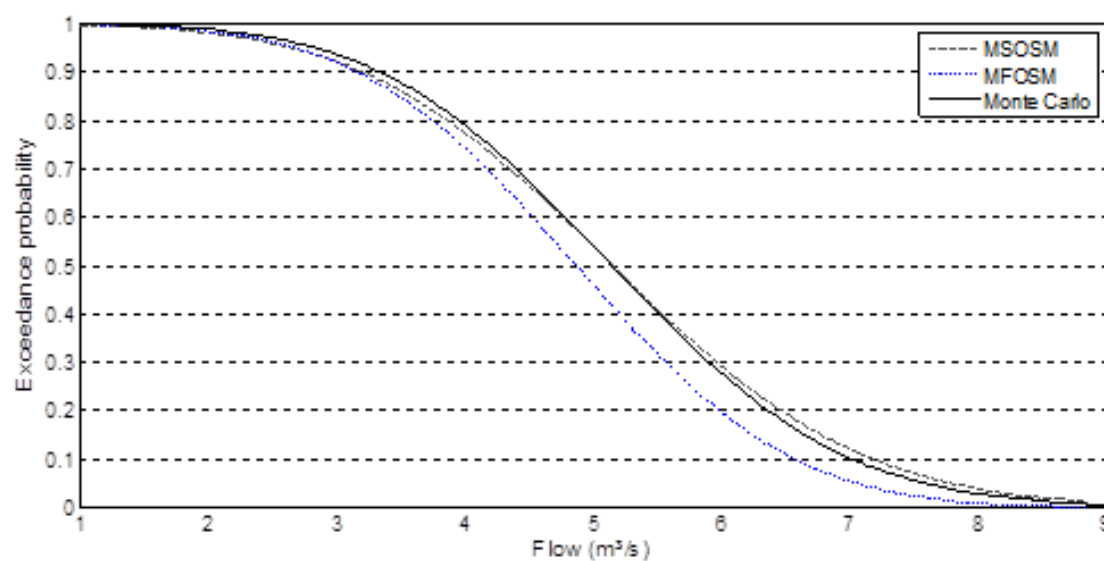


Figure 9.7 Comparaison des valeurs des EP sur le TR102 pour la journée du débit minimum résultant des différentes méthodes d'analyse d'incertitude.

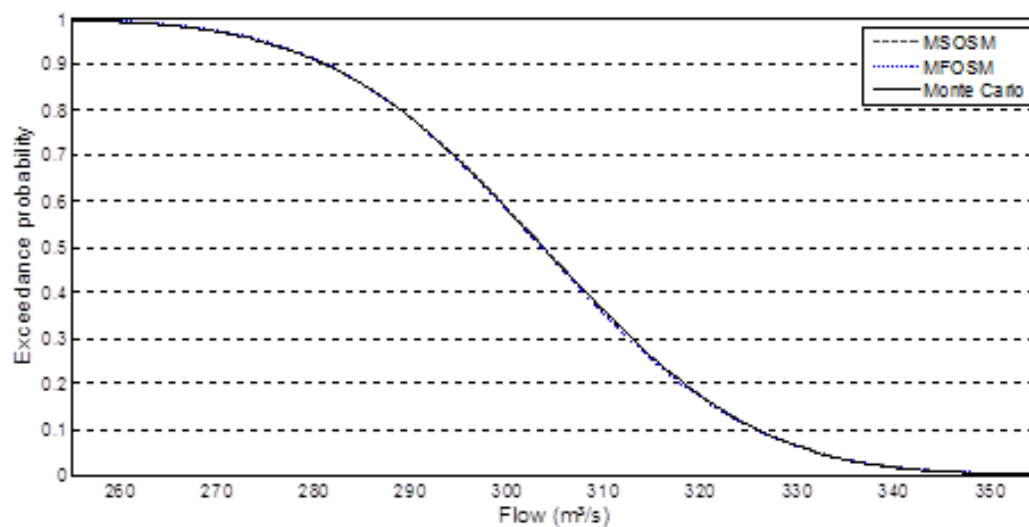


Figure 9.8 Comparaison des valeurs des EP sur le TR255 pour la journée du débit maximum résultant des différentes méthodes d'analyse d'incertitude.

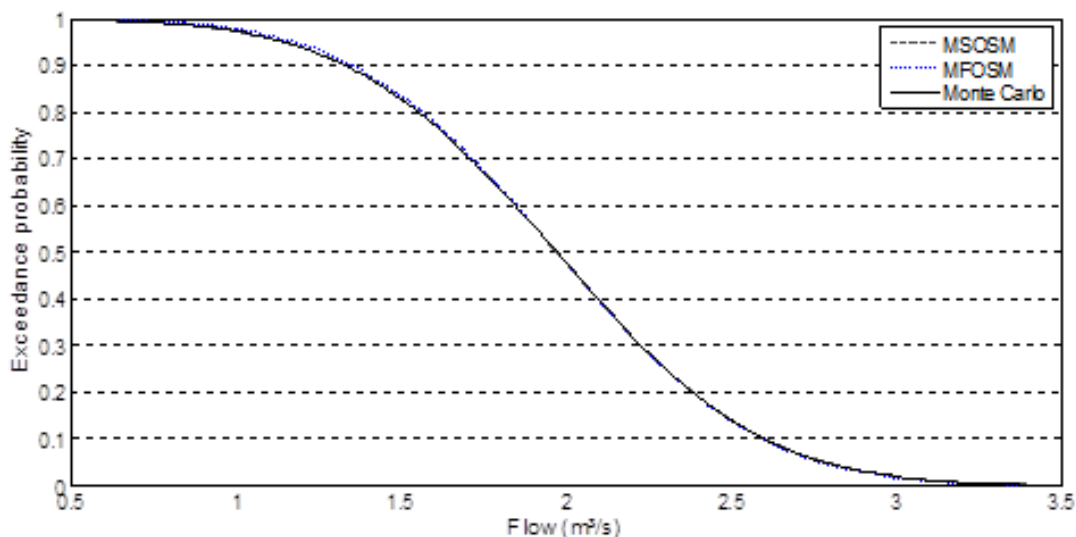


Figure 9.9 Comparaison des valeurs des EP sur le TR255 pour la journée du débit minimum résultant des différentes méthodes d'analyse d'incertitude.

9.4.1 Résultats et discussion

L'indice de sensibilité $S_j(\theta)$ donné en (8.31) a été calculé pour le bassin versant de la rivière Bécancour (l'exutoire TR1) et les trois tronçons en amont TR70, TR102 et TR255 en fixant l'incertitude du paramètre à $\Delta\theta = \pm 50\%$. Les résultats présentés dans le tableau (9.3) montrent que l'ordre des paramètres les plus influents est identique pour les deux tronçons TR70, TR102 et l'exutoire TR1, mais diffère légèrement des résultats du tronçon TR255. Ainsi, la sensibilité dépend, quelque peu, de la répartition des MH dans le bassin hydrographique.

Les MH riverains se caractérisent par des faibles valeurs de l'indice de sensibilité, en général, pour TR70 et TR102, et sont moins variables par rapport aux valeurs des indices des MH isolés, particulièrement pour TR255. En fait, pour TR1, TR70 et TR102, la variance des indices de sensibilité est supérieure à celle du tronçon TR255, pour les deux types des MH. La différence entre les indices de sensibilité peut être expliquée par la répartition des MH dans les différents tronçons ou sous-bassins versants.

Les paramètres ont été classés par ordre d'importance décroissant selon la valeur de $S_j(\theta)$. Les valeurs obtenues pour cet indice de sensibilité $S_j(\theta)$ à l'exutoire (TR1) varient de 0.007 à 0.311. En général, une sensibilité d'environ 30 % de la réponse du modèle n'est pas très élevée, mais c'est acceptable en pratique, dans le cas où seulement 12 % et 29 % de la superficie

du bassin versant est occupée et drainée par les MH (voir le tableau (9.2)). Les valeurs de l'indice de sensibilité pour TR255 varient de 0.8 % à 22 %. Le classement des paramètres est présenté dans le tableau (9.3).

Dans le cas des tronçons TR1, TR70 et TR102, le modèle est plus sensible, en ordre décroissant, aux incertitudes liées aux paramètres des MH isolés : FRAC, RAV et KSAT-BS ; et aux paramètres des MH riverains : FRAC, WETDNOR et KSAT-BS. Des valeurs de sensibilité négligeables sont assignées, dans l'ordre croissant, aux paramètres des MH isolés suivants : C-PROD et C-EV ; et aux paramètres des MH riverains suivants : WETDMAX, KSAT-BK.

Les résultats dans le cas du TR255 sont différents : le paramètre des MH riverains le plus influent est C-PROD, suivi par RAV, FRAC puis KSAT-BS, avec une variation de la sortie du modèle de 16 % à 22 %. Les paramètres qui ne contribuent guère à cette variation sont donnés, en ordre décroissant, par KSAT-BK, C-EV, WETDMAX, WETDNOR, KSAT-BS et FRAC, qui sont associés aux MH riverains, sauf pour C-EV qui est associé aux MH isolés.

Tableau 9.3 Le classement des indices de sensibilité pour chaque sous-tronçon de la rivière Bécancour

Paramètres θ	TR1 $S_j(\theta)$	Rang	TR70 $S_j(\theta)$	Rang	TR102 $S_j(\theta)$	Rang	TR255 $S_j(\theta)$	Rang
C-EV	0,097	8	0,093	8	0,087	8	0,050	9
C-PROD	0,121	7	0,122	7	0,128	7	0,221	1
FRAC	0,311	1	0,301	1	0,293	1	0,186	3
KSAT-BS	0,228	3	0,221	3	0,213	3	0,156	4
RAV	0,268	2	0,255	2	0,241	2	0,190	2
FRAC	0,190	4	0,195	4	0,191	4	0,098	5
KSAT-BK	0,007	10	0,008	10	0,007	10	0,008	10
KSAT-BS	0,155	6	0,159	5	0,156	5	0,064	6
WETDMAX	0,058	9	0,060	9	0,060	9	0,053	8
WETDNOR	0,157	5	0,159	6	0,154	6	0,059	7
Moyenne	0,159		0,157		0,153		0,108	
Écart-type	0,089		0,085		0,082		0,070	
Coefficient de variation	0,559		0,542		0,536		0,645	

CHAPITRE 10 CONCLUSION

Dans la deuxième partie de cette thèse, on a effectué une analyse de sensibilité et une analyse d'incertitude paramétrique des modules des milieux humides intégrés dans le modèle hydrologique distribué HYDROTEL. Le bassin versant de l'étude est celui de la rivière Bécancour, qui est caractérisé par plusieurs milieux humides isolés et riverains.

La modélisation probabiliste des paramètres incertains et indépendants est utilisée. En se basant sur la revue de la littérature, les paramètres sont définis par une loi de probabilité normale. La réponse du modèle est analysée statistiquement en utilisant une analyse de sensibilité locale et les méthodes d'analyse d'incertitude basées sur l'approximation du modèle par le développement en série de Taylor : MFOSM (*Mean-value first-order second-moment*) et MSOSM (*Mean-value second-order second-moment*) ainsi que la simulation de Monte-Carlo (MCS) utilisée comme méthode de comparaison et de validation des résultats obtenues par les autres méthodes. On a été en mesure de quantifier les incertitudes de la réponse du modèle (débit simulé) et d'identifier les paramètres importants qui contribuent le plus à la variabilité des débits simulés.

Concernant les différentes méthodes d'analyse d'incertitude, les probabilités de dépassement (EP) d'un seuil de débit donné, calculées par les méthodes MCS, MFOSM et MSOSM sont, en général, comparables. Les meilleures estimations des probabilités de dépassement sont obtenues par la méthode MSOSM. De là, on peut conclure que la non-linéarité de la réponse du modèle peut être légèrement significative si l'on considère uniquement les modules des milieux humides répartis sur l'ensemble du bassin versant, mais elle n'est pas du tout significative dans le cas où l'on considère le tronçon en amont TR255.

Au moyen de l'analyse de sensibilité, l'identification des paramètres les plus influents sur la réponse du modèle, permet à l'utilisateur de cibler les paramètres sur lesquels il doit s'attarder le plus lors de la mise en œuvre du modèle. Les paramètres définissant les milieux humides isolés étaient plus influents que ceux définissant les milieux humides riverains. La sensibilité du modèle semble avoir une différence significative comparativement au tronçon TR255. En effet, les débits simulés en amont du bassin versant (TR255) sont moins sensibles aux incertitudes accompagnant les paramètres, comparativement aux tronçons en aval (TR70 et TR102). Cela est dû principalement à l'effet cumulatif engendré par l'analyse des sous-bassins imbriqués (le TR1 inclut tous les tronçons, le TR70 inclut les TR102 et TR255 et le TR102 inclut le TR255). Ainsi ce qui se passe dans le TR255 est progressivement absorbé dans les tronçons en aval.

Les paramètres les plus influents à la sensibilité du modèle qui pourraient être considérés dans le cadre d'une démarche préalable du calage du modèle, sont FRAC, RAV et C-PROD associés au MH isolé. Les paramètres tels que KSAT-BK, DWETMAX et C-EV ne seront pas significatifs.

Enfin, les paramètres des milieux humides peuvent affecter significativement les variations des débits simulés et que ceux associés aux milieux humides isolés ont un impact plus important par rapport à ceux des milieux humides riverains. Ces paramètres des milieux humides sont importants et peuvent contribuer de façon positive à augmenter la performance du modèle HYDROTEL pour reproduire les caractéristiques de l'hydrogramme du bassin versant.

CHAPITRE 11 CONCLUSION GÉNÉRALE

11.1 Synthèse des travaux

La première partie (I) de cette thèse a été consacrée à : (1) la capacité du processus de renouvellement filtré considéré comme un modèle plus réaliste, à prévoir les débits un et deux jours d'avance comparativement aux processus de Poisson filtrés généralement utilisés en hydrologie, et aux processus autorégressifs d'ordre un, ces derniers s'avèrent moins performants par rapport aux processus de renouvellement filtrés, (2) l'estimation de la probabilité de dépassement des débits extrêmes ou des débits de pointes observés et, (3) trouver la distribution des débits lors du prochain signal incitant leur augmentation comparativement au processus de diffusion avec sauts et aux processus autorégressifs d'ordre un.

La deuxième partie (II) a porté sur l'analyse de sensibilité et d'incertitude paramétrique des modules de milieux humides nouvellement intégrés dans le modèle hydrologique HYDROTEL. L'évaluation de la contribution de ces modules comparativement à la capacité du modèle à reproduire les caractéristique de l'hydrogramme se fait par l'analyse d'incertitude et de sensibilité. L'analyse de sensibilité du modèle aux incertitudes associées aux paramètres des milieux humides a permis de caractériser l'impact de ces paramètres sur la réponse du modèle et d'en dégager ceux qui sont plus influents sur les débits simulés du bassin versant, ces derniers correspondent aux milieux humides isolés. Les probabilités de dépassement d'un débit donné, calculées par les méthodes MFOSM et MSOSM sont presque similaires, ce qui implique que la non-linéarité de la réponse du modèle peut être négligeable, en ne considérant que les modules des milieux humides. Enfin, il serait important de tenir compte des autres types d'incertitudes pour les prévisions hydrologiques.

11.2 Biens livrables

Au terme de l'accomplissement de ce travail un article de recherche a été publié et deux autres soumis en collaboration avec mon directeur de recherche, Mario Lefebvre. Ces articles concernent la première partie de cette thèse. Pour la deuxième partie, deux articles ont été soumis pour publication. Ces deux articles sont rédigés en collaboration avec d'autres auteurs, dans le cadre d'un stage doctoral à l'INRS-ETE.

RÉFÉRENCES

- A.N. Rousseau, S. Savary, M. Fossey, “Modélisation hydrologique des milieux humides dans les basses-terres du saint-laurent. rapport final du volet hydrologique du projet : Outils d’analyses hydrologique, économique et spatiale des services écologiques procurés par les milieux humides des basses-terres du saint-laurent : adaptations aux changements climatiques. rapport R-1436, centre eau, terre et environnement, institut national de la recherche scientifique, inrs-ete. québec, pq”, 2013.
- P. K. Andersen, “Filtered renewal processes with a two-sided impact function”, *Journal of Applied Probability*, pp. 813–821, 1979.
- , “A note on filtered markov renewal processes”, *Journal of Applied Probability*, pp. 752–756, 1981.
- M. B. Beck, “Water quality modeling : a review of the analysis of uncertainty”, *Water Resources Research*, vol. 23, no. 8, pp. 1393–1442, 1987.
- M. Bouda, A. N. Rousseau, B. Konan, P. Gagnon, et S. J. Gumiere, “Bayesian uncertainty analysis of the distributed hydrological model hydrotel”, *Journal of Hydrologic Engineering*, vol. 17, no. 9, pp. 1021–1032, 2012.
- M. Bouda, A. N. Rousseau, S. J. Gumiere, P. Gagnon, B. Konan, et R. Moussa, “Implementation of an automatic calibration procedure for hydrotel based on prior oat sensitivity and complementary identifiability analysis”, *Hydrological Processes*, vol. 28, no. 12, pp. 3947–3961, 2013.
- D. G. Cacuci, *Sensitivity and Uncertainty Analysis : Theory*. CRC Press, 2003, vol. 1.
- D. G. Cacuci, M. Ionescu-Bujor, et I. M. Navon, *Sensitivity and Uncertainty Analysis : Applications to large-scale systems*. CRC Press, 2005, vol. 2.
- R. Cont et P. Tankov, *Financial modelling with jump processes*. Chapman Hall / CRC press, 2004.
- S. Da-Veiga, “Analyse d’incertitudes et de sensibilité : Application aux modèles de cinétique chimique”, Thèse de doctorat, PhD thesis, Université Toulouse III, 2007.

- A. Dey et S. Mahadevan, “Ductile structural system reliability analysis using adaptive importance sampling”, *Structural safety*, vol. 20, no. 2, pp. 137–154, 1998.
- X. Du et A. Sudjianto, “First order saddlepoint approximation for reliability analysis”, *AIAA journal*, vol. 42, no. 6, pp. 1199–1207, 2004.
- J.-P. Fortin, R. Turcotte, S. Massicotte, R. Moussa, J. Fitzback, et J.-P. Villeneuve, “Distributed watershed model compatible with remote sensing and gis data. i : Description of model”, *Journal of Hydrologic Engineering*, vol. 6, no. 2, pp. 91–99, 2001.
- , “Distributed watershed model compatible with remote sensing and gis data. ii : Application to chaudière watershed”, *Journal of Hydrologic Engineering*, vol. 6, no. 2, pp. 100–108, 2001.
- M. Fossey, A.-N. Rousseau, F. Bensalma, S. Savary, et A. Royer, “Integrating isolated and riparian wetland modules in the PHYSITEL/HYDROTEL modelling platform : model performance and diagnosis. Article soumis 2014/07/23, révision majeure demandée 2015/01/8, version corrigée retournée 2015/02/27, accepté 2015/04/30 dans Hydrological Processes.
- A. M. Hasofer et N. C. Lind, “Exact and invariant second-moment code format”, *Journal of the Engineering Mechanics Division*, vol. 100, no. 1, pp. 111–121, 1974.
- M. Hohenbichler, S. Gollwitzer, W. Kruse, et R. Rackwitz, “New light on first-and second-order reliability methods”, *Structural Safety*, vol. 4, no. 4, pp. 267–284, 1987.
- B. Huang et X. Du, “Probabilistic uncertainty analysis by mean-value first order saddlepoint approximation”, *Reliability Engineering & System Safety*, vol. 93, no. 2, pp. 325–336, 2008.
- J. Imhof, “Computing the distribution of quadratic forms in normal variables”, *Biometrika*, pp. 419–426, 1961.
- J. Kelman, “A stochastic model for daily streamflow”, *Journal of Hydrology*, vol. 47, no. 3, pp. 235–249, 1980.
- R. W. Koch, “A stochastic streamflow model based on physical principles”, *Water Resources Research*, vol. 21, no. 4, pp. 545–553, 1985.
- F. Konecny, “On the shot-noise streamflow model and its applications”, *Stochastic Hydrology and Hydraulics*, vol. 6, no. 4, pp. 289–303, 1992.

M. Lefebvre, “Short-term hydrological forecasts using linear regression”, *Revue des Sciences de l’Eau/Journal of Water Science*, vol. 16, no. 2, pp. 255–265, 2003.

——, *Applied Stochastic Processes*. springer, 2007.

M. Lefebvre et F. Bensalma, “Modeling and forecasting river flows by means of filtered poisson processes”, *Applied Mathematical Modelling*, vol. 39, no. 1, pp. 230–243, 2015.

——, “An application of filtered renewal processes in hydrology”, *International Journal of Engineering Mathematics*, 2014, doi :10.1155/2014/593243.

M. Lefebvre et J.-L. Guilbault, “Using filtered poisson processes to model a river flow”, *Applied Mathematical Modelling*, vol. 32, no. 12, pp. 2792–2805, 2008.

H. Loaiciga et R. Leipnik, “Stochastic renewal model of low-flow streamflow sequences”, *Stochastic Hydrology and Hydraulics*, vol. 10, no. 1, pp. 65–85, 1996.

D. P. Loucks, E. Van Beek, J. R. Stedinger, J. P. Dijkman, et M. T. Villars, *Water resources systems planning and management : an introduction to methods, models and applications*. Paris : UNESCO, 2005.

S. Mahadevan, *Probability, reliability, and statistical methods in engineering design*. Wiley, 2000.

A. Mailhot et J.-P. Villeneuve, “Mean-value second-order uncertainty analysis method : application to water quality modelling”, *Advances in Water Resources*, vol. 26, no. 5, pp. 491–499, 2003.

A. H. Marcus, “Traffic noise as a filtered markov renewal process”, *Journal of Applied Probability*, pp. 377–386, 1973.

M. D. McKay, R. J. Beckman, et W. J. Conover, “Comparison of three methods for selecting values of input variables in the analysis of output from a computer code”, *Technometrics*, vol. 21, no. 2, pp. 239–245, 1979.

C. S. Melching et S. Anmangandla, “Improved first-order uncertainty method for water-quality modeling”, *Journal of Environmental Engineering*, vol. 118, no. 5, pp. 791–805, 1992.

R. C. Merton, “Option pricing when underlying stock returns are discontinuous”, *Journal of Financial Economics*, vol. 3, no. 1, pp. 125–144, 1976.

- H. Miyamoto, J. Morioka, K. Kanda, K. Michioku, T. Uotani, Y. Ohchi, et K. Aga, “A stochastic model for tree vegetation dynamics in river courses with interaction by discharge fluctuation impacts”, *Journal of Japan Society of Civil Engineers, Ser. B1 (Hydraulic Engineering)*, vol. 67, p. 1405, 2012.
- M. G. Morgan et M. Henrion, “Uncertainty : a guide to the treatment of uncertainty in quantitative policy and risk analysis”, *Uncertainty : a guide to the treatment of uncertainty in quantitative policy and risk analysis*, 1990.
- E. Parzen, *Stochastic processes*. SIAM, 1999, vol. 24.
- S. Peng et X. Zhao, “A review of uncertainty methods employed in water quality modeling”, dans *Environmental Science and Information Application Technology, 2009. ESIAT 2009. International Conference on*, vol. 2. IEEE, 2009, pp. 32–35.
- M. M. Putko, A. C. Taylor, P. A. Newman, et L. L. Green, “Approach for input uncertainty propagation and robust design in cfd using sensitivity derivatives”, *Journal of fluids engineering*, vol. 124, no. 1, pp. 60–69, 2002.
- R. Rackwitz, *Practical Probabalistic Approach to Design : By R. Rackwitz*. Technical University of Munich, Institut fur Bauingenieurwesen III, 1976.
- A. Rousseau, J. Fortin, R. Turcotte, A. Royer, S. Savary, F. Quévy, P. Noël, et C. Paniconi, “Physitel, a specialized gis for supporting the implementation of distributed hydrological models”, *Water News-Official Magazine of the Canadian Water Resources Association*, vol. 31, no. 1, pp. 18–20, 2011.
- A. Saltelli, K. Chan, E. M. Scott *et al.*, *Sensitivity analysis*. John Wiley & Sons, New York, 2000, vol. 134.
- A. Saltelli, S. Tarantola, F. Campolongo, et M. Ratto, *Sensitivity analysis in practice : a guide to assessing scientific models*. John Wiley & Sons, New York, 2004.
- R. Turcotte, J.-P. Fortin, A. Rousseau, S. Massicotte, et J.-P. Villeneuve, “Determination of the drainage structure of a watershed using a digital elevation model and a digital river and lake network”, *Journal of Hydrology*, vol. 240, no. 3, pp. 225–242, 2001.
- R. Turcotte, A. N. Rousseau, J.-P. Fortin, et J.-P. Villeneuve, “A process-oriented, multiple-objective calibration strategy accounting for model structure”, *Water Science and Application*, vol. 6, pp. 153–163, 2003.

R. Turcotte, L. Fortin, V. Fortin, J.-P. Fortin, et J.-P. Villeneuve, “Operational analysis of the spatial distribution and the temporal evolution of the snowpack water equivalent in southern québec, canada”, *Nordic Hydrology*, vol. 38, no. 3, pp. 211–234, 2007.

L. Wang, D. Beeson, et G. Wiggs, “Efficient and accurate point estimate method for moments and probability distribution estimation”, dans *Proc. of the 10th AIAA/ISSMO MAO Conference, Albany, New York, USA*, 2004.

G. Weiss, “Shot noise models for the generation of synthetic streamflow data”, *Water Resources Research*, vol. 13, no. 1, pp. 101–108, 1977.

A. Yagouti, I. Abi-Zeid, T. B. Ouarda, et B. Bobée, “Revue de processus ponctuels et synthèse de tests statistiques pour le choix d’un type de processus”, *Revue des sciences de l’eau/Journal of Water Science*, vol. 14, no. 3, pp. 323–361, 2001.

Y.-J. Yin, Y. Li, et W. M. Bulleit, “Stochastic modeling of snow loads using a filtered poisson process”, *Journal of Cold Regions Engineering*, vol. 25, no. 1, pp. 16–36, 2010.

Y.-G. Zhao et A. H. Ang, “System reliability assessment by method of moments”, *Journal of Structural Engineering*, vol. 129, no. 10, pp. 1341–1349, 2003.

Y.-G. Zhao et T. Ono, “A general procedure for first/second-order reliability method (form/sorm)”, *Structural Safety*, vol. 21, no. 2, pp. 95–112, 1999.